Transcription-driven evolution and its implications for antimicrobial resistance

Mark Nasef Ragheb

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Houra Merrikh, Chair

David Sherman

Samuel Miller

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Transcription-driven evolution and its implications for antimicrobial resistance

Mark Nasef Ragheb

Chair of the Supervisory Committee:
Professor Houra Merrikh
Department of Biochemistry (Vanderbilt University)

Evolution guides all biological phenomena, including the rapid rise of antimicrobial resistance (AMR) in bacterial pathogens. Canonical approaches to combat AMR development, such as the development of new antibiotics have largely failed. Therefore, novel approaches to curb the rise of AMR development are sorely needed. In the first part of this thesis, I will describe work showing that the Mfd translocase protein promotes the rapid evolution of AMR. I find that Mfd promotes evolution and mutagenesis in divergent bacterial pathogens and to diverse classes of antibiotics. In addition, I find that Mfd directly promotes drug-resistance conferring mutations as well as the rise of hypermutator cells, which rapidly acquire resistance. I also find that Mfd requires interactions with RNA polymerase (RNAP) and the nucleotide excision repair (NER) protein UvrA to promote the rapid evolution of resistance. We propose that Mfd may be an ideal drug target to help inhibit the evolution of antibiotic resistance.

In the second part of this thesis, I further characterize the *in vivo* function of Mfd, a long-standing enigma in the field. I identify Mfd as a critical RNAP co-factor *in vivo* and find that this activity is critical for regulating transcription at sites of highly structured regulatory RNAs. I find that Mfd's activity at regions expressing structured, regulatory RNAs is essential for cellular

viability. Lastly, I find that Mfd promotes mutagenesis at sites containing structured RNAs, linking Mfd-mediated mutagenesis to the evolution of structured RNAs. The work presented in this thesis furthers our understanding of how Mfd promotes AMR development as well as our understanding of how this protein functions *in vivo* to regulate transcription and influence evolution.

Table of Contents

Acknowledgements

First, many thanks to my advisor, Houra Merrikh, for her encouragement and infectious passion for science. The last four years have truly flown by, and I am humbled that you have invested your time and energy in training me as a scientist. I am grateful for the opportunity to have learned so much from my time with you. I will cherish these lessons as I move on to my next journey in science.

To the countless people in the Merrikh lab that have helped train me as a scientist. First, thanks to Chris Merrikh and Kevin Lang for your patient and thoughtful advice on all things science over the last four years. To Moe Thomason for helping me bounce interesting ideas off you, for immense help with writing and figure making, and for helping me stay organized and focused through the tough times. Lastly, to Ariana Samadpour for your constant encouragement and patient demeanor, helping bring sanity and calm to the challenges of science.

To all the teachers and mentors throughout the years that supported me, fought for me, and believed in my potential. I have not forgotten your dedication.

To my wonderful classmates at UW MSTP. You are all incredible, and your support over the last six years (and the upcoming two years) is immeasurable. Special thanks to Stephanie, Gabe, David, Alec (and Ted), and Rini for always being there to share in all the highs and all the lows of a brutal and winding path. A special thanks to my MCB cohort as well, specifically Dylan Udy, for your wonderful pick-and-roll game and unwavering enthusiasm for the University of Oregon. May your knees be replenished.

To Sarah- I am so grateful to have you in my life. You are kind, supportive, incredibly talented, and undoubtedly clever. I constantly look forward to spending time with you, and I cannot wait to see what the future holds.

To my dearest and closest friends (you all know who you are). Through our worldly travels, frequent mishaps, and occasional tears of joy and happiness, I have become who I am. All of you push me to be the best person I can be, and I am in constant awe of all your successes and accomplishments. I am mostly amazed that through all the successes, every one of you retains your humility, grace, and empathy. For the sake of humanity, I hope all of you reproduce and help populate this earth.

Lastly, to Mom and Dad. There is little I can really say other than I love you both very much. You have sacrificed everything for me and I cannot be grateful enough. The journey we have been through together is one only we can share and only we can know. It is a special one. And it will continue.

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Publications

The work in this thesis is from the following publications and manuscripts in preparation:

(1) Ragheb, M.N. and Merrikh, H. The enigmatic role of Mfd in replication-transcription conflicts in bacteria. Submitted to *DNA Repair*.

(2) Ragheb, M.N., Thomason, M.K., Hsu, C., Nugent, P., Gage, J., Samadpour, A.N., Kariisa, A. Merrikh, C.N., Miller, S.I., Sherman, D.R., Merrikh, H. (2019) Inhibiting the Evolution of Antibiotic Resistance. *Mol Cell*.

(3) Ragheb, M.N., Tran, S., Merrikh, C.N, Merrikh, H. Mfd is a critical RNAP co-factor that links structured RNAs to evolution. In preparation for *Cell*

Chapter 1

Introduction

1.1 Evolution and genetic mutation

In 1859, Charles Darwin published *On the Origin of Species*, bringing forward the concept that populations evolve through the process of **natural selection**[1]. The theory of natural selection is beautifully simple and profound. Within a species, individuals carry different, *heritable* traits. Individuals who carry beneficial traits outcompete others, resulting in their differential survival and reproduction, ultimately leading to the evolution of species (Figure 1.1). The laws of natural selection govern all life forms, from bacteria to humans.

At the time, the basis for heritability, a critical tenet of natural selection, was unknown. Ultimately, Gregor Mendel's discoveries on the pattern of genetic inheritance and later discoveries proving that deoxyribonucleic acid (DNA) is the molecular basis for inheritance[2] led to the modern synthesis of evolutionary biology and genetics. We now understand that the process of evolution consists of two main drivers- **genetic mutation** and natural selection.

Popular understanding of molecular evolution posits that **mutations occur randomly** on DNA, **in the absence of natural selection**. These mutations lead to genetic variation in populations and natural selection *acts upon* this variation, leading to the process of evolution. The seminal experiments showing that mutations arise without regard for environmental selection were performed by Salvador Luria and Max Delbrück. Luria and Delbrück showed that in bacteria, genetic resistance to T1 phage occurred spontaneously, *before* exposure to the virus[3].

The authors therefore concluded that mutations *do not* arise as a result of their environment, but rather that pre-existing mutations are selected upon by the environment.

Towards a more holistic understanding of mutation and evolution

In many instances, mutations arise randomly without respect to the phenotypic consequence. However, advances in our understanding of DNA mutagenesis have led to the discovery that the generation of mutation is not random[4–10]. The rate of mutation is controlled both **temporally** (i.e. the mutation rate is plastic with respect to time) and **spatially** (i.e. certain regions of the genome are more mutagenic). Whether or not the spatial and temporal mechanisms of mutagenesis discuss below are truly "directed" or "adaptive" remains controversial (for a review on this controversy, see [11]). In other words, debate remains regarding whether the primary function of these mechanisms is to promote mutation or if it is simply a necessary byproduct of the pathway. Nonetheless, it is undeniable that such mechanisms occur.

Below I will discuss the ways in which the generation of mutations is both temporally and spatially regulated. Much of our understanding on the mechanisms of DNA mutagenesis come from bacterial studies, but it is critical to keep in mind that many of the mechanisms and principles learned from these studies are relevant across all domains of life.

1.2 Mechanisms of mutagenesis

Mutation occurs when a change is made to the underlying genetic code. Mutations can manifest as a single nucleotide substitutions (e.g. cytosine → thymine), a single nucleotide insertion or deletion, or as a larger mutation such as a chromosomal inversion or copy number variation[12].

This discussion will largely focus on nucleotide substitutions. These mutations changes arise through diverse means but can largely be categorized into two groups: DNA replication errors and mutations caused via chemical modifications.

DNA replication fidelity and mutagenesis

Errors during DNA replication is one critical way by which mutations arise. In most bacteria, DNA replication initiates from a single origin of replication and proceeds bidirectionally until the replication machinery reaches the termination site[13]. Replication is carried out by the bacterial replisome, consisting of a complex of proteins, with the DNA polymerase III enzyme (Pol III), directly synthesizing DNA[14]. DNA replication is a remarkably faithful process, with the error rate being as low as 10^{-9}-10^{-10} mutations per base pair)[15].

The fidelity of replication is maintained through three mechanisms. First, fidelity is maintained through the intrinsic specificity of DNA pol III for incorporating the correct nucleotide during replication elongation, which has an error rate of roughly 10^{-5} per nucleotide per generation[16]. Second, Pol III carries proofreading ability. In many bacteria, including in *Escherichia coli* and *Bacillus subtilis*, this is mediated by the proofreading subunit, ε, of the Pol III holoenzyme[17]. The proofreading subunit contains a 3' → 5' exonuclease activity[18], which allows for removal of misincorporated bases. The exonuclease activity of ε increases replication fidelity by roughly 100-fold[17]. Not all bacterial organisms encode a separate subunit that performs replication proofreading. For example, the main replicative polymerase in *Mycobacterium tuberculosis*, DnaE1, contains its own intrinsic proofreading capacity[19].

Third, post-replicative mechanisms of DNA repair exist. In most organisms this is largely mediated by the mismatch repair (MMR) pathway. In *E. coli*, the MutS homodimer recognizes mismatched bases on DNA, leading to a conformational change in the protein that facilitates interaction with the MutL endonuclease. The interaction between MutS and MutL activates the MutH protein, leading to the MutSHL complex to bind DNA[20]. In *E. coli*, the MutSHL complex slides along the DNA until it reaches the nearest methylation site. The absence of methylation on newly synthesized DNA is how the *E. coli* MutSHL complex ensures that MMR specifically repairs the newly replicated strand[21]. MutH nicks DNA near the hemimethylated site and an exonuclease enzyme (RecJ, ExoVII if the nick is at the 5' end of the mismatch or ExoI if it is on the 3' end of the mismatch[22]) is recruited to create a single-stranded (ss) DNA gap from the hemimethylated site until it reaches the original mismatch and removes it. The ssDNA gap is then subsequently resynthesized by DNA pol III. The MMR system is critical for genome maintenance and is highly conserved across all domains of life. Mutations in the analogous pathway in humans leads to minisatellite instability and predisposes individuals to cancer[23]. Overall, MMR increases the fidelity of DNA replication in bacteria by roughly ~100 fold[24].

These three mechanisms ensure that DNA replication fidelity is maintained at the previously cited value of 10^{-9}-10^{-10} mutations per basepair per generation (Figure 1.2).

Environmental mutations: reactive oxygen species

Like any biomolecule, DNA is prone to chemical modifications, and at some frequency these modifications cause permanent changes to the genetic code. While a vast array of DNA mutagens exists, I will focus on reactive oxygen species (ROS) as one well-studied class of

mutagens, as it is one of the largest sources of DNA damage. During aerobic cellular respiration, superoxide ($O_2^{\cdot -}$), hydrogen peroxide (H_2O_2) and hydroxyl radicals ($OH^{\cdot}$) form, which can each produce oxidized DNA species[25]. The most prevalent base modification that occurs due to ROS is 7,8-dihydro-8-oxoguanine (8-oxoG). If a guanine from the free nucleotide pool is oxidized, the resulting 8-oxoG species can mispair with adenine at a high frequency, ultimately leading to an A:T → C:G transversion. Alternatively, direct oxidation of guanine on the DNA base pair can occur, leading to an 8-oxoG:C pair. A new round of DNA synthesis can misincorporate adenine across from 8-oxoG leading to a G:C → T:A transversion[26] (Figure 1.3). Highly conserved DNA repair systems exist to help cells minimize the impact of ROS on genomic mutation. The oxidized guanine (GO) system is comprised of three enzymes in *E. coli* - *mutT*, *mutM*, and *mutY*, that sanitize the oxidized guanine pool (*mutT*), help coordinate removal misincorporated 8-oxo-G (*mutM*), or help remove mismatched bases resulting from DNA oxidation (*mutY*)[27] (Figure 1.3-includes further details on this pathway). Defects in these systems lead to significant increase in the mutation rate of cells[24]. Interestingly, conditions that induce metabolic flux and consequently increase endogenous ROS, such as sublethal concentrations of antibiotics in bacteria, is thought to directly induce mutagenesis[28]. Such work underscores the concept that exposure to stressful toxins (both endogenous and exogenous) can promote mutagenesis over the course of a cell's lifetime.

When during a cell's lifetime and where in its genome a nucleotide substitution may occur is highly heterogenous. In the following section, I will discuss how such errors are controlled both temporally and spatially.

1.3 Temporal control of mutagenesis

There are times during the life cycle of an organism when it faces exogenous stress. In order to survive, organisms must adapt to these stressors. In bacteria, it has been proposed that an important adaptation under stress is the modulation of mutation rate[10]. Various ways to temporally regulate mutagenesis have been described.

Adaptive mutagenesis

Under different types of nutrient limitations bacteria can modulate their mutation rates. Seminal work by Cairns et al. proposed that under certain conditions, mutations arise *in response* to the environmental conditions[8]. In this work, the authors found that when *E. coli* auxotrophs incapable of fermenting lactose (Lac^-) grew on minimal media containing lactose, mutants accumulated that allowed for growth on lactose (Lac^+)[8]. Critically, the presence of lactose in the media *facilitated* the accumulation of Lac^+ revertants. The authors concluded from this finding that the mutations were adaptive to the presence of the environment and *not* random, as proposed by Luria and Delbrück. While the concept remains provocative, over the last decades mechanisms by which cells regulate their mutation rate have been discovered.

The SOS response

One well studied means by which bacteria alter mutagenesis in response to exogenous stress is through DNA damage sensing via the SOS response[29]. In the SOS pathway, the LexA repressor protein binds to and negatively regulates the expression of a diverse set of genes involved in DNA repair and DNA mutagenesis[30]. Upon sensing DNA damage, the RecA protein filaments around this damaged site and functions as a LexA protease, de-repressing the inhibition of LexA-

bound genes and inducing their expression. In *E. coli,* this includes the induction of low-fidelity DNA polymerases- Pol II, Pol IV, and Pol V[31]. Unlike DNA Pol III, these polymerases perform translesion (TLS) DNA synthesis (i.e. replication past large DNA lesions)[32,33], and critically they also carry a mutation rate as high as 1 in every 100 nucleotides[34]. Reports show that TLS polymerases are critical for bacterial adaptation in various contexts, ranging from stationary-phase mutagenesis (see below) to antibiotic resistance[35–37].

Starvation-induced and stationary phase mutagenesis

In addition to DNA damage, nutrient limitation and general stress can also ratchet up mutation rates in bacteria[38]. In *E. coli* the alternative sigma factor RpoS plays a critical role in facilitating mutagenesis. When *E. coli* faces resource limitation, as in the case of nutrient starvation or growth in stationary phase, the amount of active RpoS increases, which in turns increases the rate of mutation. RpoS increases mutagenesis through at least two described mechanisms- (1) through the induction of TLS polymerases[39], as previously described in the SOS response, and (2) through down-regulation of the MMR repair pathway[40,41]. In *B. subtilis*, stationary-phase mutagenesis promotes reversion of auxotrophic mutants, allowing for growth under amino acid starvation. However, this did not appear to be dependent upon the SOS pathway nor an alternative sigma factor, but rather on genes that induce pathways of genetic competence[42]. Additionally, stationary-phase mutations seem to be promoted by the Mfd translocase protein[43] (Mfd-mediated mutagenesis will be discussed in more detail below).

Hypermutation

Like all biological processes, the rate of mutation itself is subject to the process of evolution. Stressful environments can lead to the selection of increased mutation rates, or hypermutation, in bacteria. The extent of hypermutation can be as little as two-fold (seen in clinical strains of *Mycobacterium tuberculosis*)[44] or as large as ~100 to ~1000-fold (seen in experimental evolution studies in *Salmonella typhimurium*[45] and in clinical strains of *Pseudomonas aeruginosa*[46]). The mechanisms by which hypermutation occur are largely through inactivating mutations in processes that regulate DNA fidelity. For example, mutations that inactive the mismatch repair (MMR) pathway, which recognizes and excises errors made in newly synthesized DNA, increase mutation rates ~10-100-fold[46], while mutations that inhibit the proofreading capacity of DNA polymerase III can alter rates by ~100-1000-fold[45,47]. It is important to note that hypermutation as an adaptive process differs from the aforementioned mechanisms of stress-induced mutagenesis: hypermutation is an evolutionary selection that acts on the population, whereas stress-induced mutagenesis is a hard-wired, mechanistic response to exogenous conditions that occurs at the individual level. Nonetheless, both processes are useful means by which organisms adapt to changes in their environment.

In summary, the rate of mutation is not static throughout the life of a cell. Diverse mechanisms exist that facilitate genetic adaption under environmental stress and they likely play an important role in ensuring the survival of populations under natural selection.

1.4 Spatial control of mutagenesis

The likelihood of a mutation arising at any given nucleotide position in a genome is not equal. Regions of a genome that have a higher probability of mutating are termed "hotspots" in the genome. Mutational hotpots occur through a wide array of mechanisms and some of these will be discussed below.

Nucleotide substitutions and sequence context

DNA consists of four different nucleotides- adenine (A), guanine (G), thymine (T), and cytosine (C). In a mutational event, any given nucleotide can be replaced by one of the three others (e.g. T→A, T → C, T→G), resulting in 12 possible nucleotide substitutions (i.e. four different nucleotides can be replaced by three others, and 3x4= 12). Mutations can be categorized into two different types of substitutions- *transitions* and *transversions*[48]. Transversions are defined as a change from a purine to a pyrimidine (and vise-versa) while a transition is defined as a purine <-> purine change (e.g. C→T) or pyrimidine <-> pyrimidine change (e.g. A→G). While 8 of the possible 12 nucleotide substitutions are transversions, transitions mutations are more likely to occur[49]. The higher frequency of observed transition mutations is thought to be due, at least in part, to the biochemical properties of complementary base pairing[50]. Transition mutations are caused by nucleotide deamination as well as tautomerization, while transversion mutations can be caused by different sources of DNA damage, including ROS[51]. Recent advances in sequencing technologies have allowed for a more detailed understanding of differences in base substitutions. Work by Lee, et al. using a novel maximum-depth sequencing approach to identify rare mutations found a higher prevalence of transitions as expected, but unexpectedly found a high prevalence of C→A transversion mutations[52]. The authors speculate that perhaps

ribonucleotide adenines (as opposed to deoxyribonucleotides) are misincorporated at high frequency into the genome[52]. Additionally, these C→A transversion mutations occur more frequently when there is a neighboring C, suggesting that surrounding sequence context plays a factor in the mutation. This work highlights the fact that much is still not understood about the nature of mutations in the genome.

DNA structures and nucleotide substitutions

In addition to different probabilities of base substitutions occurring at different bases, the sequence context surrounding any given nucleotide can alter the rate of mutation. Repetitive DNA elements- encompassing dinucleotide repeats, inverted repeats, minisatellites, microsatellites, amongst other elements- contribute significantly to genetic variability (reviewed in [53]). Many of these elements form alternative DNA structures, which are genetically unstable and promote large genomic alterations such as copy number variation and chromosomal rearrangements. In addition, increasing evidence shows that these sequence elements also promote nucleotide substitutions[54]. In *Saccharomyces cerevisiae*, varying the length of GAA repeat expansions can significantly increase the rate of nucleotide substitutions, up to one kilobase away from the site of the repeat element[55,56]. Such effects are not specific to trinucleotide repeat expansions- other sequence structures, such as inverted repeat elements in DNA, also promote an increased rate of substitution[57].

Multiple mechanisms exist that are thought to promote base substitutions at unstable DNA elements and alternative DNA structures (reviewed in [54]). The stalling of RNA polymerase (RNAP) at short repeats and the subsequent recruitment of nucleotide excision repair (NER)

machinery (discussed more extensively in sections below) has been invoked as a possible contributor to mutagenesis[58]. In this model, excessive endonucleolytic activity by NER factors generates DNA end resection and subsequently gap-filling via error-prone eukaryotic TLS polymerases promotes point mutations. Indeed, genetic evidence supports the model that TLS polymerases contribute significantly to mutations at these sites[59–61]. A second model by which DNA elements promote point mutations is through the phenomenon of break induced replication (BIR). In BIR, large DNA elements induce reversal of the DNA replication fork, leading to a DNA structure analogous to a Holliday junction[62]. Resolution of this structure can lead to the formation of a one sided double-stranded break (DSB) which is repaired via the BIR pathway[63] (Figure 1.4). Replication via BIR is highly mutagenic[64] and various factors are thought to contribute to this mutagenesis (reviewed in [65]). In particular, the frequent dissociation of DNA replication, the possible contribution of low fidelity polymerases, and the conservative nature of DNA replication in this pathway[66] (which subsequently occludes the activity of MMR[54]), all may contribute to mutagenesis during BIR[65].

Recent evidence implicates genomic instability induced at repeat elements as critical for evolution of stickleback fish[67], highlighting the importance of alternative DNA structures in influencing rates of evolution. While much of the work regarding DNA elements and mutations are from mammalian studies, DNA sequence context is also a major determinant of mutagenesis in bacteria[68,69]. Given that many of the pathways and mechanisms that promote instability at these sites are conserved across all domains of life, it is plausible that DNA elements and the alternative structures they form may promote evolution of bacterial species.

In summary, the rate of mutation can vary significant from nucleotide to nucleotide across the genome. As is the case with many mechanisms of mutagenesis, whether or not these differences are directed or simply a consequence of other factors remains an interesting point of discussion.

Transcription associated mutagenesis (TAM)

While it is well appreciated the mutations arise through the process of replication, it is less appreciated that transcription also promotes mutagenesis[70], and provides another means by which mutations are spatially regulated. TAM can be divided into three, non-mutually exclusive (and in fact overlapping), categories: replication-transcription conflicts, R-loops, and the translocase protein, Mfd. I discuss all three below.

R-loops

One mechanism by which transcription promotes mutations is through R-loops. R-loops are transcription dependent, three-stranded nucleic acid structures consisting of nascent mRNA annealed to its complementary coding strand along with a displaced single DNA strand[71]. It is thought that R-loops promote mutations via the presence of exposed single-stranded DNA, known to be sensitive to cellular mutagens such as reactive oxidative and nitrogen species[70,72]. By definition, the accumulation of R-loops occurs at transcribed sites, but other factors can additionally enhance R-loop formation, such as high GC content and head-on replication-transcription conflicts[73,74] (discussed below). Because R-loops threaten genome integrity and promote genomic instability, organisms have evolved critical R-loop resolution factors termed RNase H enzymes[75]. These enzymes degrade the RNA strand of an R-loop, thereby facilitating the reannealing of the DNA duplex. Interestingly, studies in *E. coli* and *B. subtilis* show that

deletion and overexpression of these RNase H enzymes increase and decrease mutations, respectively[73,76], providing evidence that modulating R-loop levels alters mutation rates.

Replication-transcription conflicts

DNA replication and transcription must occur in a timely and accurate fashion in all organisms. Individually, each one of these tasks is a remarkable undertaking. It is even more striking to consider that replication and transcription utilize the same DNA template, yet, the machineries responsible for these respective processes must not impede each other's function. While cells are remarkably efficient at coordinating replication and transcription, at some frequency, there are unavoidable encounters between the replication machinery (the replisome) and RNA polymerase (RNAP)[77–80]. The encounters between these machineries, termed replication-transcription conflicts, come in two different flavors: head-on (if a gene is encoded for on the lagging stand of replication) or co-directional (if a gene is encoded for on the leading strand of replication) (Figure 1.5). Conflicts can lead to various detrimental outcomes, including replication fork stalling[81] and double-stranded DNA breaks (DSBs)[82,83]. One of the other major consequences of replication-transcription conflicts is DNA mutagenesis[84–86]. Head-on conflicts are more mutagenic than co-directional ones (although increased transcription increases mutagenesis in both orientations)[84,86]. The mechanism by which head-on conflicts promotes increased mutagenesis is dependent upon both Mfd[85] (discussed below) as well as R-loop formation, as overexpression of RNase H enzymes in *B. subtilis* decreased mutations at head-on conflict regions[73]. The link between the contributions of Mfd and R-loops to mutations at head-on conflicts remains to be fully elucidated. One potential unifying mechanism is that R-loop

formation may be induced by Mfd[76] (perhaps by promoting transcription elongation), although this relationship has not been studied in the context of conflicts.

Mfd and mutagenesis

Mfd is a highly conserved translocase protein that interacts directly with RNAP and is thought to be involved in multiple cellular processes, which will be discussed at length below. Mfd promotes mutagenesis in various contexts. In *B. subtilis*, a deletion of Mfd reduced the number of revertant mutants of multiple transcribed reporter genes in the context of stationary-phase mutagenesis[43,87]. Starvation conditions in *E. coli* induces mutagenesis through the formation of double-stranded breaks (DSBs), which are generated by R-loop formation in an Mfd-dependent manner[76].

Mfd also promotes mutagenesis at head-on replication-transcription conflicts. Million Weaver et al. discovered that the enhanced mutation rate of head-on reporter genes in *B. subtilis* was dependent on Mfd, as deletion of Mfd reduced mutation rates of the head-on genes studied[85]. Additionally, the authors found an epistatic relationship between Mfd and other factors of the transcription-coupled repair pathway (TCR) and proposed a model by which TCR activity at head-on conflicts promotes error-prone DNA repair[85]. The mutagenic potential of Mfd is likely to be highly conserved, as various studies show that Mfd promotes antibiotic resistance development in different bacterial pathogens[88–90]. While it is clear that Mfd promotes mutagenesis, the detailed mechanism by which this process occurs remains a mystery.

Summary

Mutation is the substance for evolution and understanding the ways in which mutations arise is critical to better understanding how organisms adapt to their environment. In this section, I have discussed the many ways in which mutations arise, largely in bacterial genomes.

It is critical to note that temporal and spatial regulation of mutation is simply one of many overlapping ways to categorize mechanisms of mutagenesis, and such demarcations are not always clean distinctions. For example, TAM is a mechanism by which organisms spatially regulate mutagenesis, but it also a way in which organisms can potentially temporally regulate mutagenesis (e.g. head-on conflicts could increase mutagenesis both spatially as well temporally, as timing of head-on gene expression could be considered a temporal regulation of mutagenesis[91]). Nonetheless, the regulation of mutagenesis may play a critical role in the evolution of species.

1.5 The mutagenic translocase factor, Mfd

The Mfd translocase protein promotes TAM in various contexts. Here, I will provide a thorough background on Mfd, as this protein will be extensively discussed in subsequent chapters. I will begin by providing an overview of Mfd's biochemical functions and proposed *in vivo* functions. I will address some of the outstanding questions regarding this protein's controversial and enigmatic activities, and I will discuss its proposed role in mutagenesis in various contexts.

An overview of Mfd

The research on Mfd spans over fifty years, yet much is still left to learn about this highly enigmatic protein. Mfd is classically described as a DNA repair protein that links transcription

and the nucleotide excision repair (NER) repair pathway via the transcription-coupled repair (TCR) pathway[92,93], leading to preferential repair of the transcribed strand of DNA[94,95] (Figure 1.6- which includes a more detailed description of the TCR pathway).

Much of what we know about Mfd's function is through biochemical experiments. Mfd is a superfamily 2 (SF2) helicase protein with ATPase dependent translocase activity on DNA[96]. The 130 kDA Mfd protein consists of eight unique structural domains that can be further categorized into three unique biochemical domains: the UvrB homology module (consisting of domains D1a/D2/D1b), the RNAP interacting domain (RID) (consisting of domain D4), and the translocation domain (consisting of D5/D6)[97]. The UvrB homology module and RID domain are the contact points for interactions with the UvrA protein of NER and the beta-subunit of RNA polymerase, RpoB. These interactions, in particular the interaction with RpoB, are required for all currently described *in vivo* functions of Mfd. Specifically, in the context of TCR, Mfd recognizes **stalled RNAP** (canonically due to a bulky DNA lesion such as a cyclopyrimidine dimer), is loaded onto DNA and subsequently utilizes its ATPase and forward translocase activity to displace stalled RNAP from DNA[98,99], essentially pushing RNAP away from the offending lesion. Translocation of RNAP exposes the offending lesion to NER proteins, which Mfd recruits to the site of damage via binding to UvrA[97]. For Mfd to bind UvrA, a conformational change is induced upon RNAP binding such that the UvrB homology module, which is normally obstructed by the D7 domain of the protein, is accessible for binding to UvrA[100]. One major mystery regarding the biochemical activity of Mfd is precisely how it discriminated between stalled RNAP and normally translocating RNAP. One possible

explanation is that the reduced rate of RNAP elongation at a stall site provides a kinetic window for Mfd to bind to RNAP and perform its biochemical functions[101].

DNA repair and Mfd

While there is an abundance of biochemical details regarding Mfd's activity, the importance of these functions *in vivo* remains controversial. Mfd's *in vivo* functions are likely important as the protein is highly conserved across all domains of life and is constitutively expressed, at least in *E. coli* and in *B. subtilis*. Moreover, mutations of the functional homolog of Mfd in humans, CSB, leads to the Cockayne Syndrome, a debilitating neurodegenerative disease[102,103]. As previously mentioned, historically, Mfd is considered the central link between transcription and NER. What remains unexplained, however, are data from multiple studies showing little to no sensitivity to ultraviolet (UV) radiation, one of the main DNA-damaging agents repaired via NER. Recent work has implicated other proteins outside of Mfd in coordinating TCR. In *E. coli*, the transcription elongation factor NusA is thought to promote Mfd-independent TCR. NusA physically interacts with UvrA, and unlike Mfd, cells deleted of NusA are sensitive to DNA lesions that block RNAP elongation[104]. It was subsequently proposed that in *E. coli*, the UvrD helicase was largely responsible for TCR. In this model, UvrD facilitates TCR by pushing RNAP backwards on DNA (i.e. promoting RNAP backtracking) to expose the DNA lesion, in contrast to the previously discussed forward translocase model of Mfd-dependent TCR. Because UvrD and Mfd would theoretically oppose either others translocase activity, it was proposed that "anti-backtracking" factors, such as Mfd may actually *interfere* with DNA repair[105]. Indeed, in a different study, the same group showed the overexpression of Mfd can actually *sensitize* cells to

DNA damage suggesting that at least in some cases, Mfd interferes with DNA repair[106]. Overall, these data point to the possibility that Mfd's main cellular function may be outside of repair.

Mfd rescues arrested RNAP

More recent work indicates that Mfd has a broader cellular role and provided greater mechanistic insights into its function. Various *in vitro* and *in vivo* studies now show that Mfd utilizes its translocase activity to move extensively backtracked RNAP (that is, RNAP that has moved backwards on the DNA template) forward as well as to remove arrested RNAPs at a wide variety of obstacles that are *not* DNA lesions, ranging from protein roadblocks to sites of nucleotide starvation[101,107–111]. Recent *in vitro* work shows that Mfd translocates autonomously (independent of RNAP) upon binding to DNA[101]. Consequently, Mfd is capable of recognizing RNAP arrest downstream of the initial site where it was loaded. Mfd's function on arrested RNAP depends on the severity of the roadblock as it will either rescue severely backtracked RNAPs or promote transcription termination if the obstacle cannot be overcome[101]. Additionally, single-molecule microscopy imaging of Mfd reveals that it binds to DNA *in vivo* in the absence of exogenous stressors[112], corroborating *in vitro* findings and providing further evidence that Mfd may function as a fundamental RNAP processivity factor outside of RNAP-blocking lesion such as UV irradiation. Evidence in mammals also suggests that CSB promotes transcription elongation[113,114], in addition to its role in DNA repair through TCR. It is therefore possible that Mfd's most critical cellular function is to help clear stalled RNAPs from DNA in an efficient manner. However, the endogenous roadblocks Mfd helps rescue and the cellular consequences of an Mfd deletion under these conditions remain unknown.

Is Mfd a critical replication-transcription conflict resolution factor?

Given that the replication fork is a significant roadblock to RNAP processivity, it is logical to examine the role of Mfd in reducing the severity of replication-transcription conflicts. Intriguingly, the answer seems complex. Trautinger et al. provided the initial genetic evidence for the role of Mfd in replication-transcription conflicts. The authors showed that a deletion of Mfd (in addition to other factors that deal with stalled RNAP) sensitized cells to UV damage when cells were additionally deficient in DNA repair[115]. This effect was eliminated in the presence of a mutant RNAP (*rpo*35*), which destabilizes RNAP elongation[115]. The authors propose a model whereby UV stalled RNAP complexes promote genomic instability upon encounters with replication. In this context, RNAP-associated factors such as Mfd seem to be important for maintaining genomic stability.

Pomerantz and O'Donnell undertook the first mechanistic study directly looking at the role of Mfd in conflicts. These authors recapitulated the bacterial replisome *in vitro* and loaded it head-on to an arrested RNAP complex on a linear dsDNA fragment. They found that replication across the linear fragment was inhibited by the presence of the head-on RNAP. The addition of Mfd into the system significantly reduced the replication stall generated by RNAP[116]. The authors concluded from their findings that Mfd activity helps promote replication through head-on transcription units, likely by displacing RNAP from DNA (Figure 1.7). Previous work from the same group suggested that Mfd is not necessary for resolution of co-directional conflicts due to the ability of replication to reinitiate using an mRNA primer[117].

Shortly after these findings, Dutta, et al. developed an *in vivo* system to investigate genomic instability at replication-transcription conflict regions as well as the importance of various RNAP associated factors in mitigating DNA damage at conflict sites. Using a plasmid based system, the authors measured the presence of DSBs at head-on and co-directional conflict sites and found that they were dependent on RNAP backtracking[118]. While conflicts induced DSBs in both orientations, Mfd was able to reduce DSBs but only when the conflict was co-directional[118] (Figure 1.7). These findings contrast with the findings of Pomerantz and O'Donnell, which implicated Mfd in helping replisome progression through head-on but not co-directional conflict regions.

Why Mfd is unable to reduce DSBs at head-on conflicts remains unclear. Perhaps *in vivo,* Mfd in unable to access or remove RNAP at severe head-on conflicts. Mfd is known to require access to at least 25 basepairs of DNA upstream of RNAP to bind to the protein[110]. It is possible that at a highly expressed head-on gene, tightly arrayed RNAPs block access of Mfd (Figure 1.8). Additionally, recent work from Lang, et al. show that R-loops preferentially form at head-on conflict sites[73]. Since R-loops likely form upstream of elongating RNAPs, they may also inhibit Mfd accessibility (Figure 1.8). Lastly, it is critical to note that the effect of Mfd in relieving DSBs at co-directional conflicts is in the presence of a strong transcription arrest via an engineered protein roadblock. It is not clear whether Mfd relieves DSBs at other co-directional conflict sites (e.g. pause sites, endogenous roadblocks).

Many conflict resolution factors are either essential proteins or become essential upon induction of a head-on conflict. Although Mfd can reduce DSBs at co-directional conflicts and help

replisome progression *in vitro*, evidence suggests that Mfd does not help cells survive the consequences of conflicts. Boubakri, et al. and De Septenville, et al. show that inversion of rRNA genes to the head-on orientation in *E. coli* require the activity of the DinG, Rep, and UvrD helicases[119] as well as the RecBC double-stranded end processing enzymes[82] for cell viability. However, deletion of Mfd, either alone or in conjunction with the aforementioned helicases, does not affect cell viability[82]. Mfd is one of various RNAP accessary factors (e.g. rho, greA/greB) that may be capable of dealing with the consequences of stalled RNAPs at replication-transcription conflicts. Perhaps this redundancy in the system explains the lack of cell viability phenotypes associated with an Mfd deletion in the presence of severe replication-transcription conflicts.

Mfd and transcriptional termination

It has been proposed that Mfd functions to help transcription terminate in bacteria. In bacteria, transcription termination is mediated through two main mechanisms- intrinsic termination or factor-dependent termination (Reviewed in [120]). During intrinsic termination, a sequence specific signal promotes destabilization and consequent dissociation of RNAP. Specifically, a uridine-rich sequence promotes RNAP pausing, allowing for subsequent formation of a GC rich terminator hairpin structure within the RNAP exit channel. This promotes weakening of the RNA-DNA hybrid contacts within the transcription bubble, facilitating release of RNAP from both the DNA and RNA. In factor-dependent termination, the hexameric helicase protein Rho protein binds to the nascent RNA transcript, translocates on RNA and associates with RNAP. Rho utilizes its motor force to promote release of RNAP from DNA and RNA by driving RNAP forward on DNA and collapsing the transcription bubble[121,122]. Mfd, much like Rho, promotes

forward translocation of RNAP in a similar faction. It has been proposed that in the presence of an RNAP roadblock, the movement of Mfd on RNAP generates a torsional force that ultimately leads to the collapse of the transcription bubble and release of RNAP and its nascent transcript from DNA[121]. On the other hand, unlike Rho, which binds to and translocase on the RNA transcript itself and seems to have some sequence specificity in its activity[123,124], Mfd is not known to terminate transcription in a regulated fashion (i.e. there are no known Mfd-dependent termination sites). Therefore, where in the genome Mfd would function to terminate transcription remains unknown.

1.6 Evolution in action: the case of bacterial drug-resistance

The process of evolution, and the previously discussed mechanisms of mutagenesis that help drive this process, is infinite in its biological consequences. I will now focus my discussion on how evolution impacts biology to the specific case of **antimicrobial resistance (AMR) development**. AMR development is a critical public health concern, as the rapid evolution of resistant organisms continues plague much of the world's poorest populations. A mechanistic understanding of the process of AMR will help promote novel approaches for combating this urgent crisis. In this overview, I will describe the major public challenges of AMR, the mechanisms by which drug resistance arises, and discuss potential unexplored sources of bacterial mutagenesis *in vivo*.

The crisis of AMR

While accurate estimates of AMR-related deaths are often difficult to attain, it is unquestionable that the numbers are stark. Current estimates suggest that in 2014 nearly one million deaths could

be attributed to AMR[125], in addition to significant morbidity and a massive economic toll. Unsurprisingly, the world's poorest nations face the overwhelming burden of AMR. The factors that contribute to the ongoing rise of AMR are widespread- ranging from poor antibiotic stewardship, to a lack of investment in new antibiotics, and to insufficient health systems in many parts of the word. While the solution to the AMR challenge is multidisciplinary, novel scientific approaches to combating AMR are sorely needed and will have significant global consequences.

Mechanisms of AMR acquisition

The idea that antimicrobial resistance could be selected as a consequence of drug exposure was described as soon as the first clinically used antibiotic, penicillin, was discovered[126]. Thereafter, the rapid development of AMR has shortly followed the deployment of every major clinically used antibiotic (Table 1.1). Bacterial drug resistance occurs in two main ways: horizontal gene transfer (HGT) and chromosomal mutation. The relative importance of these two depends on the bacteria as well as the antibiotic. Certain organisms rely heavily on chromosomal mutations to generate resistance- this is particularly true of *Mycobacterium tuberculosis (Mtb)*, where only chromosomal mutations have been observed to drive antibiotic resistance[127]. Additionally, resistance mechanisms to certain classes of drugs, such as rifamycins and fluroquinolones, arise solely through chromosomal mutations[128,129].

Mechanisms of mutations and AMR

We now understand that a critical mechanistic driver of resistance development is DNA mutation. Errors during DNA replication are the best understood source of chromosomal mutations, but as

previously discussed, non-replicative mechanisms of mutagenesis also exist, such as environmental mutations and TAM[9,70,72,130]. In order for errors introduced during DNA replication to drive antibiotic resistance, DNA replication must be active. However, for the intracellular pathogens *Mtb* and *S. typhimurium*, data suggests that DNA replication *in vivo* is significantly reduced[131,132]. Consequently, non-replicative drivers of mutations may be of critical consequence in the development of antibiotic resistance during clinical infections in these species. However, the source of DNA mutation in the context of an infection is largely unknown. To this date, the only significant study looking at DNA mutagenesis *in vivo* was performed by Ford et al. In this study, the authors infected cynomolgus macaques with *Mtb* and using whole-genome sequencing revealed that the mutational signature of *Mtb* within the host was that of oxidative damage[133]. Further research into how various mechanisms of mutagenesis help generate the genetic diversity needed for the development of antibiotic resistance may enhance our ability to combat this challenge.

Conclusions and perspectives

AMR development poses a severe threat to human health, and its continued rise will continue to strain health systems and societies worldwide. A fundamental understanding of how mutations arise during the context of clinical infections is sorely lacking. However, advances in our understanding of the diverse mechanisms of mutagenesis that exist is likely to be relevant to the context of bacterial evolution. Therefore, applying these findings to the scourge of clinical drug-resistance is the next frontier into gaining insight into the nature of AMR development and more critically may lead to alternative therapeutic strategies to combat this intractable challenge.

1.7 Figures

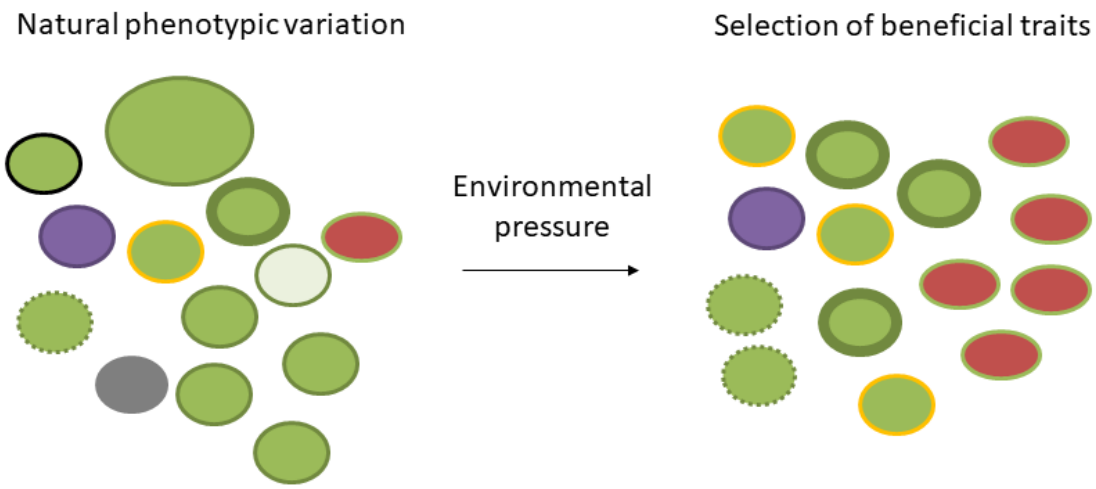

Figure 1.1 A schematic of natural selection
As proposed by Charles Darwin, the natural diversity that exists within a species can be acted on through the process of natural selection. In the presence of any environmental pressure (e.g antibiotic exposure or starvation conditions in bacteria), those within the species that carry beneficial traits will have a relative survival benefit. Because the beneficial traits are hereditary, they will be propagated over time, ultimately changing the composition of the population.

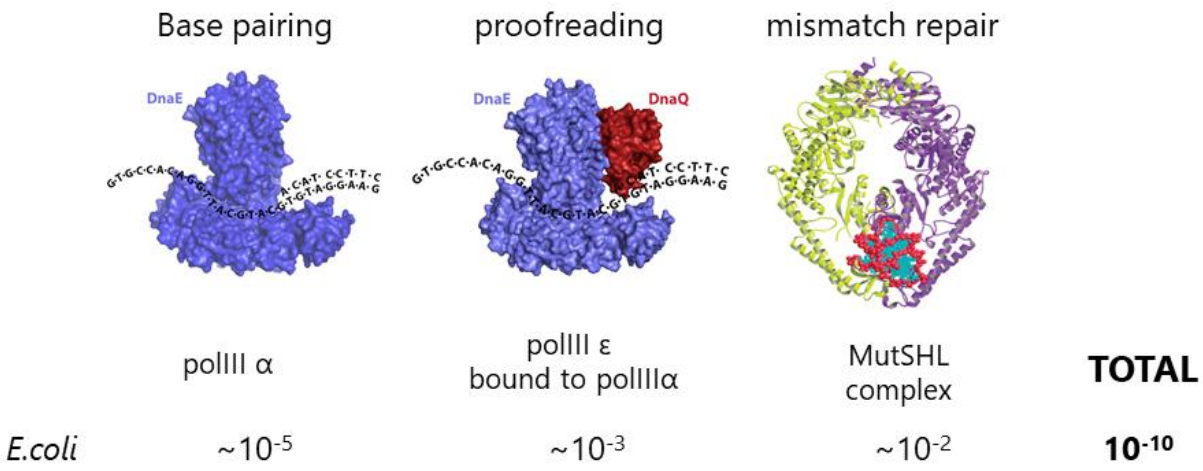

Figure 1.2 Mechanisms of DNA replication fidelity in *E. coli*
The error rate of DNA replication in most bacterial species hovers around 10^{-9} to 10^{-10} mutations per basepair per generation. In *E. coli*, this is achieved through three mechanisms. (Left) The intrinsic specificity polIIIα, the major replicative polymerase, is such that an error is made once every 10^{-5} basepairs. The exonuclease subunit of the DNA Pol III (polIIIε- center) increases replication fidelity by ~1000 fold through its ability to excise misincorporated nucleotides. Lastly, the mismatch repair system (Right- structure from PDB accession code: 1EWQ[134]) recognizes post-replicative errors in newly synthesized DNA, increasing replication fidelity by ~100 fold.

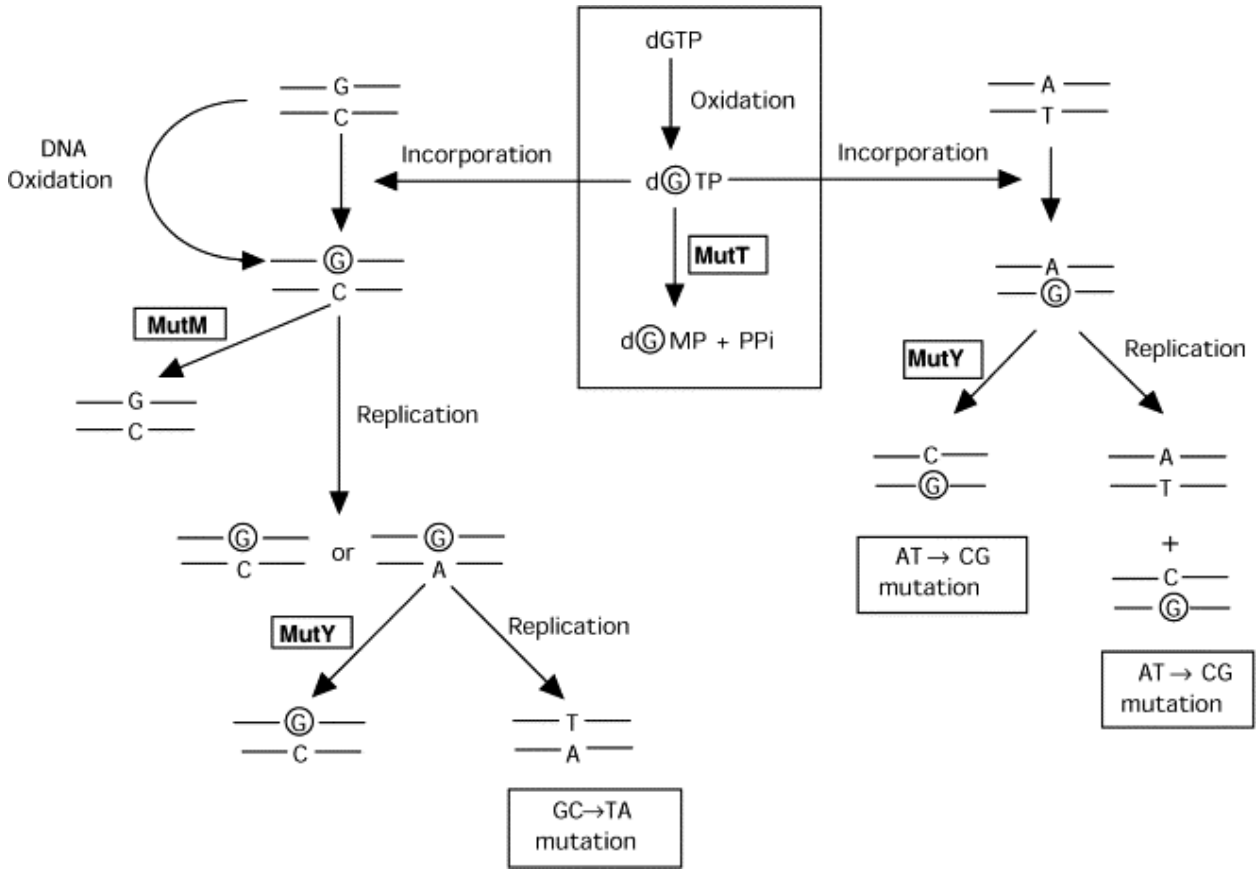

Figure 1.3 The guanine oxidation (GO) DNA repair system in *E. coli*
The *in vivo* repair of oxidized guanines in *E. coli* is largely mediated through three different proteins. The MutT protein can recognize oxidized dGTP (top center box) and catalyze the formation of oxidized dGMP, which is unable to be incorporated into replicating DNA. MutM recognizes oxidized guanines misincorporated into DNA, and help catalyzes its removal (top left image). Lastly, MutY (right image and bottom left image) can recognize A:G or A:8-oxoG mismatches on DNA and effectively remove the mismatched adenine. This figure is modified from Fowler, et al.[135]

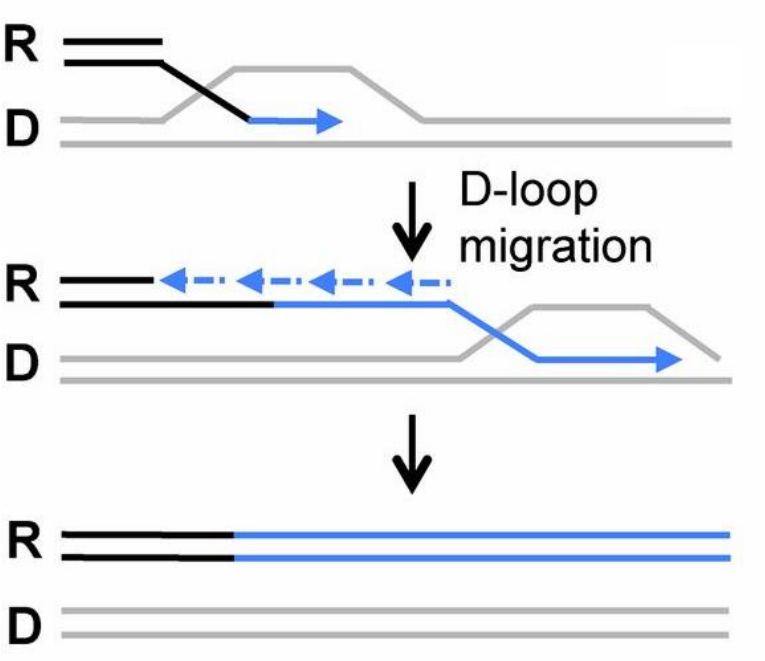

Figure 1.4 The mechanism of break induced replication (BIR)
During BIR, a single-ended break (which can occur due to fork reversal) is resolved via conservative DNA replication. In this process, the D-loop formed during strand invasion migrates while lagging strand synthesis is generated from the newly synthesized template, resulting in conservative replication. Various factors (discuss in the BIR section), contribute to mutagenesis in this pathway[65]. This figure is adapted from Donnianni and Symington[136].

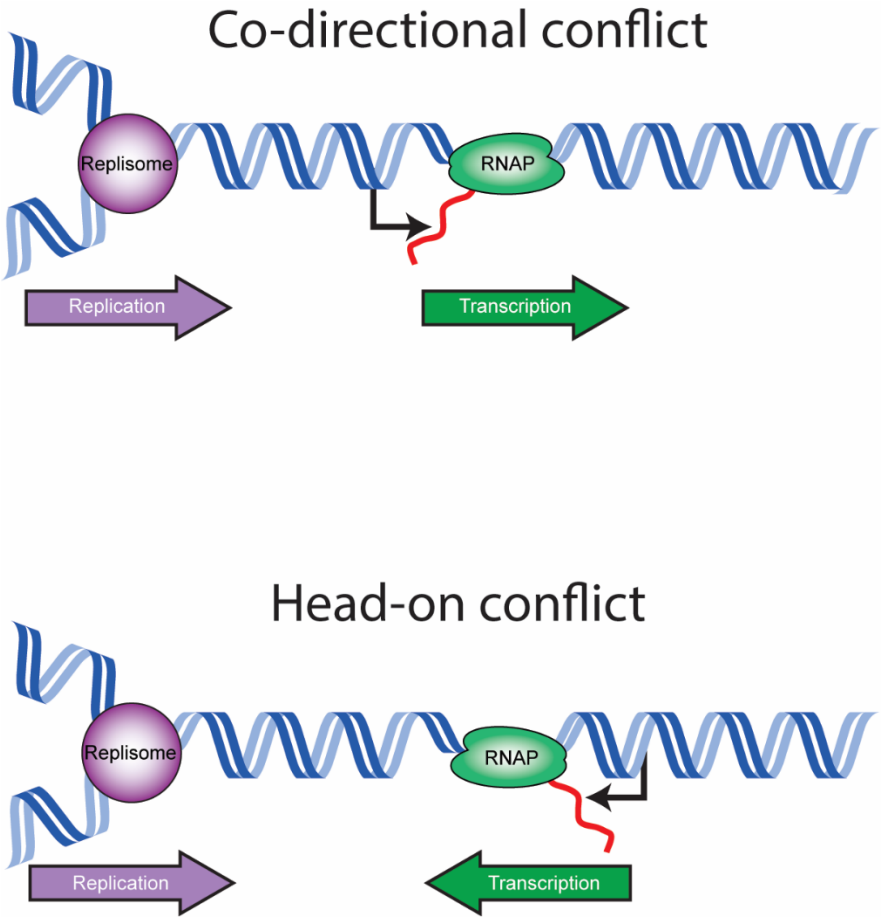

Figure 1.5 Replication-transcription conflicts in bacteria
Replication and transcription utilize the same DNA template, and these processes are not temporally segregated. This leads to encounters between the replication and transcription machineries on DNA. These replication-transcription conflicts occur in two different ways: if transcription is on the leading strand, the two machineries will be co-oriented (top image- co-directional conflict). Alternatively, if transcription is on the lagging strand, these two machineries will be processing towards each other (bottom image- head-on conflicts). Head-on conflicts are more deleterious and promote higher rates of mutation.

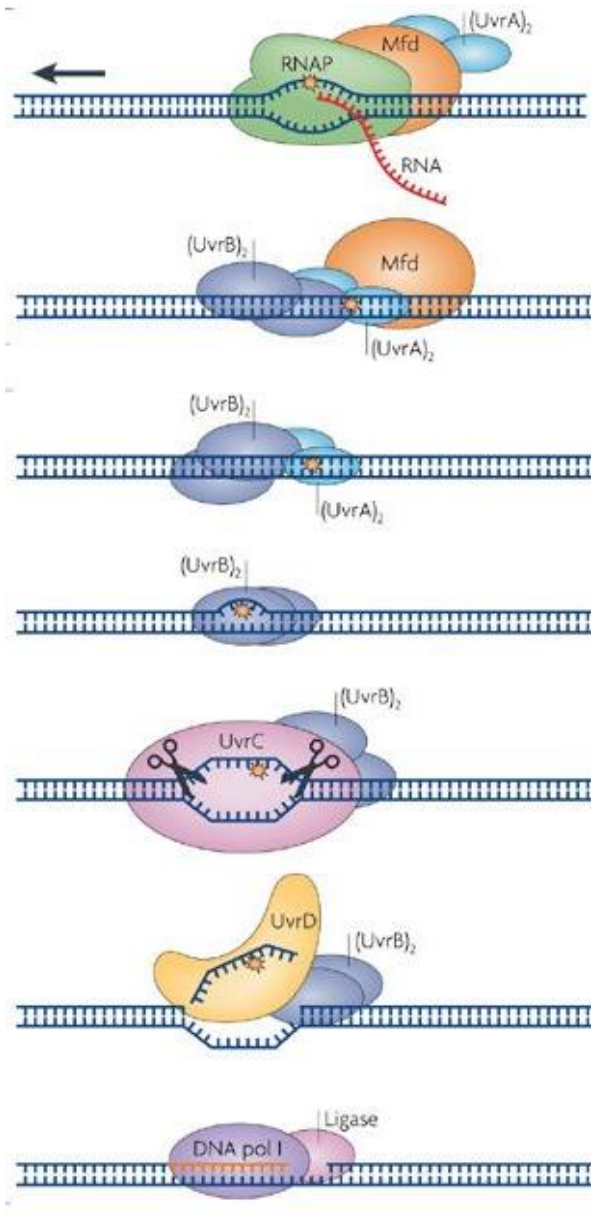

Figure 1.6 The canonical transcription-coupled repair (TCR) pathway in bacteria
In this pathway, RNAP (show in green), stalls at a bulky DNA legion (e.g. cyclopirymidine dimer). This stalling invokes binding of the Mfd protein to RNAP, inducing forward translocation of the complex to expose the offending lesion to the NER pathway. Mfd directly interacts with and recruited the UvrA homodimer, which subsequently interacts with UvrB to form the $UvrA_2UvrB_2$ complex. This complex recognizes distortions in the DNA helix induced by the lesion. Once this distortion is recognized, the UvrC endonuclease cleaves a 13 nucleotide gap surrounding the lesion site, which is removed by the UvrD helicase. The gap is ultimately synthesized by a DNA polymerase. This figured is adapted from Hanawalt and Spivak [92].

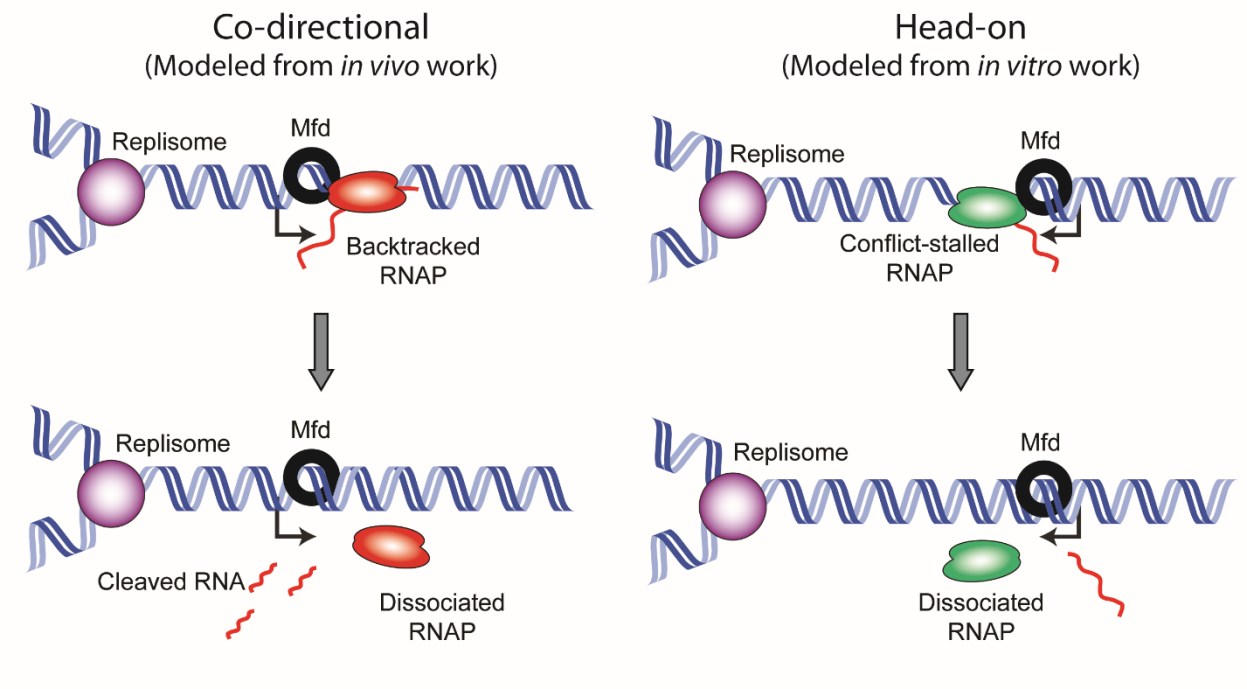

Figure 1.7 Models of conflict resolution by Mfd
The left panel shows Mfd in resolution of a co-directional conflict, cause by a backtracked RNAP. This model is based on the observations by Dutta et al.[118], *in vivo,* in *E. coli*. Mfd does not appear to be important for conflict resolution, at least in the context of backtracked RNAP, *in vivo,* The right panel shows a model for resolution of head-on conflicts, which is based on work by Pomerantz and O'Donnell[116]. The *in vitro* work placed a single RNAP ahead of the fork, and in that context, Mfd helped resolve the conflict, but only in the head-on orientation.

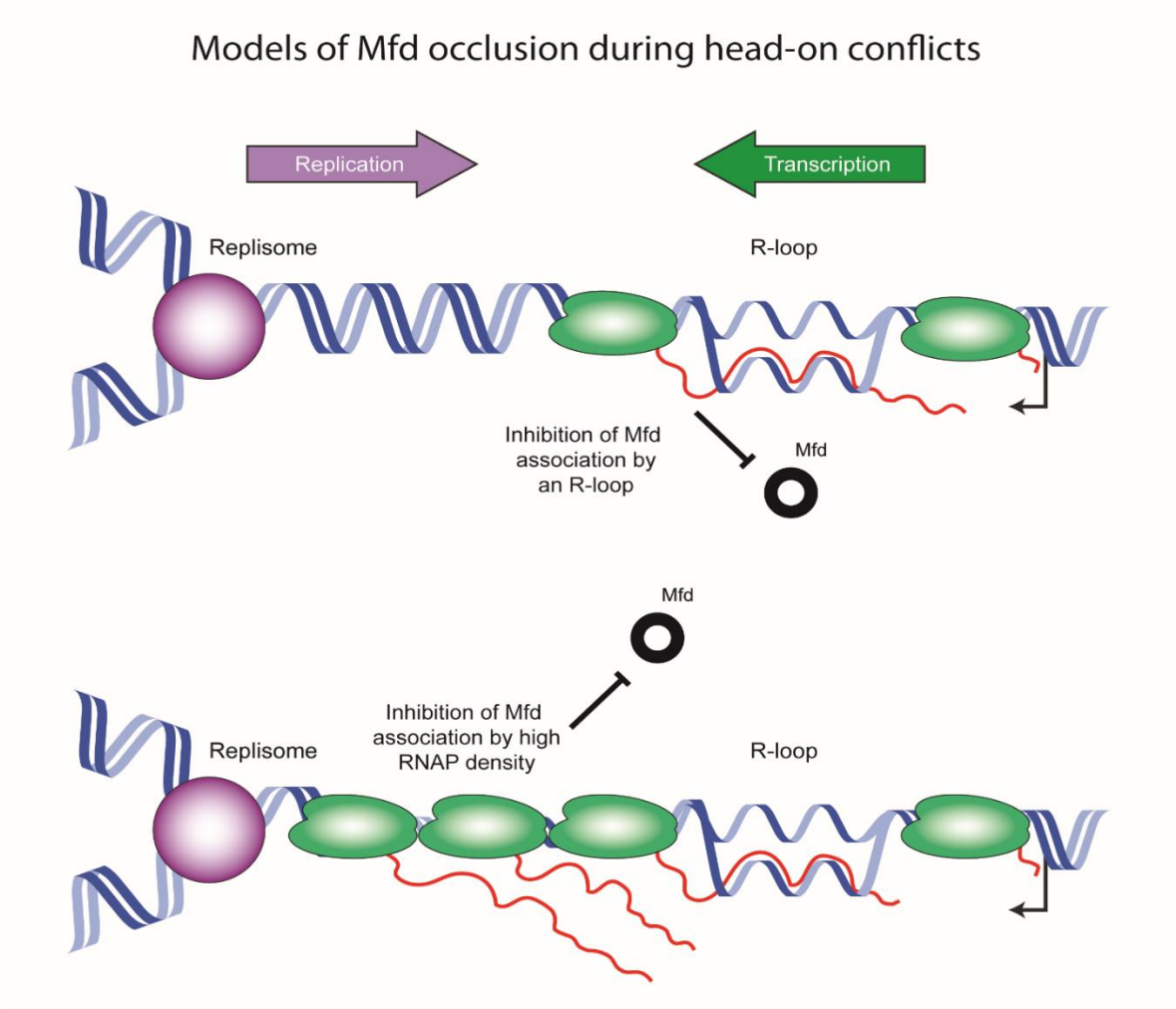

Figure 1.8. Models for why Mfd may not resolve head-on conflicts in the *in vivo* context
There are two fundamental differences between the *in vitro* work, where Mfd was observed to resolve head-on conflicts, and what takes place *in vivo* at regions of head-on transcription. Lang et al.[73] have shown that R-loops accumulate at head-on conflict regions. It is possible that the R-loops prevent Mfd binding to RNAP and occlude Mfd from the conflict region (Top). However, in contrast to the in vitro set up, the majority of genes *in vivo* are transcribed by more than a single RNAP. It is very likely that the 25bp gap needed for Mfd to sit on DNA is not available in the head-on conflict regions when RNAP density is high at a given gene (Bottom).

Antibiotic and year of deployment	Year of resistance identified
penicillin-1943	1943
tetracycline-1950	1959
methicillin- 1960	1962
gentamicin- 1967	1979
vancomycin -1972	1987
levofloxacin- 1996	1997
linezolid- 2000	2001
ceftaroline- 2010	2011
bedaquiline- 2012	2012

Table 1.1 Antibiotic resistance timeline
List of representative antibiotics, the years which they were clinically deployed (left), and the years which resistance was identified (right).

Chapter 2

Inhibiting the evolution of antibiotic resistance

2.1 Summary

Efforts to battle anti-microbial resistance (AMR) are generally focused on developing novel antibiotics. However, history shows that resistance arises regardless of the nature or potency of new drugs. Here, we propose and provide evidence for an alternate strategy to resolve this problem: inhibiting evolution. We determined that the DNA translocase Mfd is an "evolvability factor" that promotes mutagenesis and is required for rapid resistance development to all antibiotics tested, across highly divergent bacterial species. Importantly, hypermutator alleles that accelerate AMR development did not arise without Mfd, at least during evolution of trimethoprim resistance. We also show that Mfd's role in AMR development depends on its interactions with the RNA polymerase subunit RpoB and the nucleotide excision repair protein UvrA. Our findings suggest that AMR development can be inhibited through inactivation of evolvability factors (potentially with "anti-evolution" drugs), and in particular Mfd, providing an unexplored route towards battling the AMR crisis.

2.2 Introduction

The battle between antimicrobial resistant pathogens and antibiotic therapy is an evolutionary arms race – one that we are currently losing. Consequently, antimicrobial resistance (AMR) related deaths have reached alarming rates throughout the world. Estimates suggest that at least 700,000 people die annually from drug-resistance infections; this number could rise to 10 million by 2050, far surpassing cancer as the major cause of death worldwide [125]. Most efforts to resolve AMR are geared towards the development of novel antibiotics, yet resistance has arisen to every antibiotic used in the clinic. Innovative strategies to reduce the rise of drug resistant pathogens are therefore a necessary public health concern.

For many pathogens and antibiotic classes, *de novo* mutations play a critical role in AMR development. For example, in the case of *Mycobacterium tuberculosis* (*Mtb*), the causative agent of tuberculosis (TB), AMR acquisition arises exclusively through chromosomal mutations[127]. Given the alarming global burden of TB drug resistance, in addition to the rise of chromosomally-acquired AMR in many other pathogens, reducing the mutational capacity of organisms could significantly inhibit their ability to develop AMR. This approach requires the identification and subsequent inhibition of active factors that increase mutation rates. We term these proteins "evolvability factors" given that they can promote evolution by increasing mutation rates (either directly or indirectly).

The DNA translocase protein Mfd is highly conserved across bacterial phyla, suggesting that it plays an important physiological role in cells. Like its functional analogue CSB in humans, Mfd's main function has long been thought to be in the initiation of nucleotide excision repair (NER - which repairs bulky lesions on DNA) at sites of stalled RNA polymerases (RNAP)[92]. This mechanism is referred to as transcription-coupled repair (TCR). Detailed biochemical studies have

provided insight into the various functions of Mfd, including its role in the recruitment of NER proteins to regions of stalled RNAP. Curiously though, cells lacking Mfd do not display increased sensitivity to DNA damaging agents[104–106,137,138]. Furthermore, overexpression of Mfd sensitizes cells to DNA damage[106]. Moreover, even though Mfd is canonically known to promote DNA repair, it paradoxically increases mutagenesis in certain contexts, such as at regions of replication-transcription conflicts and in stationary-phase mutagenesis[43,76,85,88,89,139]. These findings can be interpreted in at least three different (but not mutually exclusive) ways: 1) redundant TCR mechanisms exist (also proposed by Kamarthapu et al., 2016)[106], 2) Mfd may actually inhibit DNA repair in some contexts (also proposed by Pani and Nudler, 2017)[140], and 3) Mfd may promote DNA repair, but this repair is mutagenic in the absence of exogenous DNA damage (e.g. Million-Weaver et al., 2015)[85].

Mfd may have additional functions outside of TCR. Recently, Mfd was found to associate with RNAP even in the absence of exogenous DNA damage[101,112], suggesting that it may play a more general housekeeping role during transcription elongation. Furthermore, Mfd acts as an RNAP anti-backtracking factor and therefore could be critical for RNAP processivity. Mfd's anti-backtracking activity also alleviates genomic instability caused by collisions between replication and transcription elongation complexes[118].

Here, we identify Mfd as an evolvability factor, the absence of which hinders antibiotic resistance development. We show that Mfd promotes mutagenesis in bacteria both during laboratory growth and during infection of eukaryotic cells. Our experiments show that the Mfd-dependent increase in mutagenesis accelerates AMR development, and this holds true for multiple classes of antibiotics. We also find that Mfd promotes the evolution of hypermutation, one important mechanism known to lead to rapid AMR development. Importantly, our findings show

that the role of Mfd in AMR development is highly conserved across bacteria, including mulitple clinically relevant pathogens. Finally, we pinpoint critical regions of Mfd that are required for its evolvability function. Specifically, we show that the interactions of Mfd with the RNA polymerase beta subunit, RpoB, as well as the NER protein, UvrA, are required for its role in the rapid evolution of resistance to several classes of antibiotics. Altogether, these results provide evidence that blocking evolvability factors, and in particular Mfd, can inhibit resistance development in a diverse array of bacterial pathogens.

2.3 Results

Mfd is a mutagenic factor in divergent bacterial species

The role of Mfd in DNA repair has remained controversial: cells lacking Mfd are not sensitive to DNA damaging agents and previous work hints at a mutagenic role for Mfd in specific contexts. We decided to thoroughly examine Mfd's role in mutagenesis, specifically in the absence of exogenous DNA damage. We measured mutation rates, with and without Mfd, in divergent bacterial species using Luria-Delbrück fluctuation analysis[3]. We observed that strains lacking Mfd had a 2 to 5-fold decrease in mutation rates as measured by rifampicin resistance compared to wild-type (WT) strains (Figure 2.1). This decreased mutation rate was conserved between Gram-negative and Gram-positive species, including *Bacillus subtilis* and clinical isolates of *Salmonella typhimurium*[141] and *Pseudomonas aeruginosa*[142] (Figure 2.1A). These results are in contrast to Mfd's previously published anti-mutagenic properties during UV exposure[93,143].

Chromosomal mutations are the sole means by which AMR develops in *Mtb*[127]. Differences in mutation rates between clinical isolates of *Mtb* is thought to underlie AMR development[44]. Therefore, if conserved, Mfd-driven mutagenesis could have a significant impact

on the development of AMR in *Mtb*. Indeed, when we deleted the gene encoding Mfd, mutation rates in *Mtb* were reduced by roughly 2-3-fold (Figure 2.1B), as measured by resistance to three different antimicrobials frequently used to treat tuberculosis: rifampicin, ethambutol, and ciprofloxacin. This suggests that Mfd promotes mutagenesis across different resistance loci in *Mtb* and may be critical for the development of antibiotic resistance.

Mfd's mutagenic function is conserved during infection of eukaryotic cells

We wanted to determine if the mutagenic effects of Mfd are conserved in an infection model of drug resistance. For these experiments, we infected CACO-2 epithelial cells with a clinical isolate of *S. typhimurium*, and subsequently measured mutation frequency using resistance to 5-flourocytosine[144]. Interestingly, compared to the ~2-4-fold decrease observed during laboratory growth (Figure 2.1C, left), we see a ~5-fold decrease in mutagenesis in the absence of Mfd upon host cell infection (Figure 2.1C, right). These differences are not related to growth defects during infection of host cells, as there is no change in the number of colony forming units following infection (Figure 2.1D). Therefore, the effect of Mfd-mediated mutagenesis is both conserved, and potentially enhanced, during growth and replication in the host.

Mfd accelerates AMR development

We next assessed the impact of Mfd on both the kinetics and the levels of AMR development in short term evolution experiments in the Gram-negative pathogen, *S. typhimurium*. Given that the differences in mutation rates between WT and cells lacking Mfd were modest (2-5 fold), we wondered if these differences could impact the kinetics and evolution of resistance in a meaningful way. To test this model, we developed an assay that measures both metrics over

roughly 35 to 70 generations, in the presence of antibiotics (ranging from sub-inhibitory to ~16x MIC for our first time-course). We monitored the evolution of *S. typhimurium* resistance to a panel of clinically relevant antibiotics (rifampicin, phosphomycin, trimethoprim, kanamycin, and vancomycin), which act through different mechanisms and have different resistance loci. We found that resistance to all the antibiotics tested arose significantly faster and to higher levels in WT compared to the Δ*mfd* strain (Figure 2.2A-E). The difference in the median resistance levels between the two strains at the end of the *S. typhimurium* evolution experiments was 6-21-fold greater in WT compared to Δ*mfd*. These results were not specific to *S. typhimurium*: we found a ~32 fold difference in median antibiotic concentration tolerated by WT compared to Δ*mfd* cells in the highly divergent, Gram-positive bacterium, *B. subtilis* (Figure 2.2F). These findings show that Mfd promotes resistance development in diverse bacterial species.

Mfd is critical for the development of AMR in *Mtb*

Mtb is arguably the most difficult-to-treat pathogen due to AMR development. Therefore, we were interested in determining whether Mfd is responsible for the evolution of resistance in this pathogen. Therefore, we adapted our short-term evolution assays to the unique culture conditions of *Mtb*, and performed the evolution experiments using rifampicin as a representative antibiotic. The difference in median resistance to rifampicin between the two strains at the end of the experiment was striking: the median resistance level to rifampicin in WT was in some experiments up to 1000-fold higher than Δ*mfd* strains (Figure 2.3A). This difference is significantly greater than that observed for *S. typhimurium* or *B. subtilis*. Additionally, we find that by the end of our evolution assays, roughly 2/3 of our evolved WT strains were resistant above the clinical MIC breakpoint of *Mtb* to rifampicin (1mg/L) while none of the Δ*mfd* strains reached this

threshold[145]. These data suggest that, as observed in other species, Mfd is critical in the development of AMR in *Mtb,* a finding with potential clinical implications.

The evolvability function of Mfd can be cross-complemented between divergent species

The data presented above suggest that Mfd's role in AMR development is conserved across species. To test the degree of conservation, we first performed bacterial 2-hybrid assays to determine if Mfd's well-documented interaction with RpoB can be detected between *S. typhimurium* and *Mtb* proteins. We chose to test these species for our experiments because they are highly divergent. Furthermore, although minimal, these two species have the biggest difference in the Mfd sequence at the amino acid level. We found that *S. typhimurium* RpoB interacts with the *Mtb* Mfd-RNAP interaction domain (RID) (Figure 2.3B). These data suggest that the mechanism by which Mfd promotes the evolution of AMR could be conserved across these species.

We then performed cross-complementation experiments using *S. typhimurium* and *Mtb* as models (Figure 2.3C-D). We introduced a copy of the *Mtb mfd* gene into *S. typhimurium* strains lacking *mfd* and performed both mutation rate and evolution assays to rifampicin. Strikingly, the *Mtb mfd* gene fully complemented the reduced mutation rates (Figure 2.3C) as well as the delayed evolution of Δ*mfd S. typhimurium* resistance to rifampicin (Figure 2.3D). These results indicate that the mechanism facilitating the evolvability function of Mfd is highly conserved across bacterial species.

Mfd promotes evolution by increasing mutagenesis

To determine if Mfd promotes AMR development through its mutagenic properties, we used Sanger sequencing to identify mutations that arose within the known rifampicin and trimethoprim resistance genes (*rpoB*[146] and *folA*[147], respectively) during our evolution experiments. Analysis of the sequences obtained from every time point, for 12 different replicates in both *S. typhimurium* WT and Δ*mfd* strains, revealed several resistance mutations. However, the Δ*mfd* replicates consistently accumulated roughly 1/2 the number of mutations in *rpoB* and 1/3 of the number of mutations in *folA* compared to WT (Figure 2.4A-B). Importantly, we observed a significant delay in the acquisition of mutations in the Δ*mfd* strains compared to those in WT and rarely observed additional second and third mutations in the Δ*mfd* strains (Figure 2.4A-B). These data strongly suggest that Mfd promotes the evolution of resistance to antibiotics through its pro-mutagenic function, and may be critical for the acquisition of multiple mutations.

Mfd promotes the rise of hypermutators

To determine if there were any mutations outside of the resistance loci in WT compared to Δ*mfd* strains, we performed whole-genome sequencing (WGS) of 6 randomly chosen replicates from our rifampicin and trimethoprim evolution experiments. WGS of 6 WT and Δ*mfd S. typhimurium* isolates from every time point of our rifampicin and trimethoprim evolution assays confirmed that compared to our WT strain, Δ*mfd* strains accumulated significantly fewer mutations in the *rpoB* locus, as we had observed using Sanger sequencing (Figure 2.4A-B). We did not find any additional mutations outside of the *rpoB* gene in any of the evolved rifampicin resistant strains that we sequenced. In contrast, our WGS of strains evolved in trimethoprim revealed the presence of additional mutations outside of the coding region, within the putative promoter region of *folA*.

All 6 sequenced WT strains contained one of two putative promoter mutations (either 35 or 61 base-pairs upstream of the *folA* coding sequence), while only one of our sequenced Δ*mfd* isolates carried one of these mutations.

Interestingly, we found that 3 out of 6 WT sequenced trimethoprim-evolved strains contained a point mutation in the *dnaQ* gene (all strains had the same *dnaQ*(I33N) mutation), while none of the Δ*mfd* strains contained any mutations in the *dnaQ* gene (data not shown). Mutations in *dnaQ* are known to generate hypermutator phenotypes[148], so, to determine if this new allele indeed conferred a hypermutator phenotype, we performed Luria-Delbrück fluctuation analysis of an evolved WT strain before and after gaining the identified *dnaQ* mutation. We found that the mutation rate upon gaining this *dnaQ* allele was ~1000-fold higher than the ancestor strain (Figure 2.5). We subsequently performed Sanger sequencing of the *dnaQ* allele on 4 additional WT and Δ*mfd* strains and found that 2 out of 4 WT strains contained the same *dnaQ* mutation, while none of the 4 Δ*mfd* strains contained this mutation. Overall, we can estimate that roughly 50% of WT strains developed hypermutator alleles during the evolution of trimethoprim resistance, while strains lacking Mfd are restrained in developing this phenotype (we did not find a hypermutator Δ*mfd* isolate).

As expected, we found that WT *S. typhimurium* isolates carrying the *dnaQ* hypermutator allele accumulated a high number of mutations across the genome (up to 600 in some of our evolved isolates), including mutations that may confer an adaptive advantage in the presence of trimethoprim. These mutations should be examined further to discern true adaptive mutations from hitchhiker mutations. These potentially adaptive mutations are in genes previously implicated in promoting trimethoprim resistance[47], such as *aroK* (involved in the shimikate pathway for folate synthesis) (data not shown). We also found putative adaptive mutations in genes not previously

associated with trimethoprim resistance. These mutations arose in genes such as *fis* (DNA binding and regulator of replication initiation and global transcription), *pyrG* (CTP synthetase), *ygdP* (RNA pyrophosphohydrolase), and *ybgC* (Acyl-CoA thioester hydrolase), amongst many others (data not shown). Mutations in many of these genes arose in independent lineages, suggesting that they may confer adaption to trimethoprim. These findings show that in our evolution assays (as in clinical settings), the generation of hypermutation may offer an adaptive strategy to evolve high-level antibiotic resistance, and that Mfd promotes this phenomenon.

Mfd-mediated evolution requires its interaction with RNAP and UvrA

To test whether the evolvability function of Mfd depended on transcription and its conserved interaction with RpoB, we constructed an L499R mutation in the RNAP interaction domain (RID) of *S. typhimurium* Mfd. The Mfd L499R mutant was previously characterized in *E. coli* and was shown to alter Mfd-RpoB interaction without affecting Mfd's DNA binding and ATPase activity[97]. Our bacterial 2-hybrid assays confirmed that disrupting this residue in *S. typhimurium* Mfd abrogates its binding to RpoB (Figure 2.6A). We found that WT Mfd fully complements the decreased evolvability of Δ*mfd* strains, whereas the L499R point mutant cannot complement the evolution of resistance to rifampicin (Figure 2.6B-C), trimethoprim, or phosphomycin (Figure 2.7). Therefore, the interaction between Mfd and RpoB is essential for Mfd's mutagenesis and subsequent evolvability function.

We next sought to determine whether Mfd's interaction with its other known binding partner, UvrA, was also necessary for its evolvability function. In order to test this, we constructed a point mutation in the D2 domain (part of the UvrB homology module) of the Mfd protein, known to mediate binding to UvrA[97]. Bacterial 2-hybrid assays confirmed that a point mutation in the

UvrA interaction domain of *S. typhimurium* Mfd (R165A) successfully abrogated its binding to UvrA (Figure 2.6A). We expressed *S. typhimurium* Mfd containing the R165A point mutation in the Δ*mfd* background and tested the impact of this mutant on mutation rates and evolution of resistance to rifampicin or phosphomycin. The Mfd R165A mutant was unable to complement any of the phenotypes (lowered mutagenesis or restricted evolution of resistance) associated with the lack of Mfd (Figure 2.6B-C, Figure 2.7). The D2 domain of Mfd is also thought to form an interface with the C-terminus of Mfd (the D2/D7 interface) as a means of restricting Mfd-UvrA interactions in the absence of RNAP stalling[97]. While we cannot exclude the possibility that the R165A mutation alters this interaction, it is unlikely that we can attribute a disruption of the D2/D7 interface to our observed results. The major effects of abrogating D7 activity is thought to be tighter binding of Mfd to UvrA[97,100,149] yet the phenotypes of this mutant mimic those of the Δ*mfd* strains. Overall, our data suggest that Mfd promotes mutagenesis and evolution of drug-resistance through its interactions with both RpoB and UvrA.

2.4 Discussion

In this work, we assign a novel function to Mfd as an evolvability factor and demonstrate that it accelerates AMR development. We show that Mfd's evolvability function requires its evolutionarily conserved interactions with both RpoB and UvrA. Arguably, the ability to evolve is critical for bacterial survival under ever-changing environmental conditions. This is especially important in the context of pathogenesis, where escaping host immunity is essential and requires constant adaptation. Therefore, our model of Mfd as an evolvability factor could explain its high degree of conservation across phyla, especially in pathogens, which are unlikely to experience TCR activation through UV exposure. This model is consistent with data showing that a deletion

of Mfd has minor effects on DNA damage tolerance, especially compared to cells missing NER[137,138] (Figure 2.8) and that bacteria lacking Mfd are insensitive to the DNA damaging environment within macrophages (Figure 2.9). These observations suggest that Mfd is largely dispensable for the coordination of DNA repair. TCR is still an important and effective method of lesion detection and repair, however, Mfd may not be the main driver of this DNA repair mechanism. Recently, a new TCR pathway[104] driven by the helicase UvrD[105,106] was discovered, indicating that cells also harbor Mfd-independent TCR mechanisms.

How Mfd promotes antimicrobial resistance and mutagenesis is unclear. One possible explanation is that Mfd promotes mutagenic DNA repair through error-prone gap filling at sites of NER activity as previously suggested[85]. Mfd may also promote DNA repair at sites that do not contain damaged DNA, given that Mfd can associate with RNAP in the absence of exogenous DNA damage[112]. *In vitro* data showing that NER can promote gratuitous repair of undamaged DNA, leading to recurrent DNA re-synthesis, which could consequently promote mutagenesis, is also consistent with this model[150]. Alternatively, Mfd may promote mutagenesis by inhibiting the activation of other DNA repair pathways at least under "normal" growth conditions – e.g. absence of UV damage. These pathways may include Mfd-independent TCR or global NER. Given our data with the UvrA interaction mutant of Mfd, we would predict that such inhibition would be through sequestration of UvrA.

Mfd-mediated evolution may be critical in the context of host infection. During infections, bacterial replication is reduced[131,132], consequently reducing replication fork errors and possibly enhancing the relative contribution of non-replicative mutations[151]. Given that transcription is still active under these conditions, Mfd may play a critical role in promoting bacterial mutagenesis during infections. This may explain the exaggerated effects of Mfd that we observed in our

infection model. Additionally, our evolution assays, which mimic the variable antibiotic concentrations seen during clinical infections, suggest that Mfd is required for rapidly developing high-levels of drug resistance upon primary exposure to sub-inhibitory concentrations of antibiotics, which may be critical in the context of AMR development[152].

Lastly, Mfd may be even more important when multiple mutations are necessary to confer resistance, such as in the context of multi-drug resistance acquisition or in the context of compensatory mutations. Our sequencing data are consistent with this prediction given that the occurrence of second and third mutations was quite rare in the absence of Mfd. Additionally, we observe that the rise of *dnaQ* hypermutator strains is inhibited in the absence of Mfd. Hypermutation is a key strategy that bacteria use to evolve resistance in the context of infections [46,153]. However, strains containing specifically the *dnaQ* hypermutator alleles have not been identified in clinical settings. This could simply be due to the fact that few (if any) isolates from trimethoprim treated patients have been sequenced. Therefore, it is still possible that the *dnaQ* hypermutator allele is relevant in clinical settings. This question should be investigated by WGS of pathogens isolated from trimethoprim treated patients.

Given our findings, we propose that blocking evolvability factors, and in particular Mfd, could be a revolutionary strategy to address the AMR crisis. A new class of "anti-evolution" drugs that target Mfd, or other evolvability factors that promote mutagenesis may complement new antimicrobials and alleviate the problem of chromosomally acquired mutations that promote AMR. For example, LexA, which induces the SOS response upon exposure to DNA damage, has been suggested to promote AMR development through Trans-Lesion Synthesis (TLS) at replication forks[154–156]. This mechanism could also be a good target for the inhibition of AMR development. However, SOS-mediated AMR development may be distinct from the transcription-dependent

evolvability function of Mfd, which (in addition to replicating cells) could be relevant in infections where pathogens are not replicating, and/or have not been exposed to extensive DNA damage, but are transcriptionally active. Therefore, in principle, drugs that target Mfd (LexA or TLS polymerases) could be co-administered with antibiotics during treatment of infections, reducing the likelihood of resistance development at the onset of treatment. Overall, efforts to understand and target the evolutionary capacity of cells could also have wide-ranging implications outside of AMR development, from reducing cancer evolution, to limiting pathogenic diversity in the context of host immunity.

The ideas discussed here deliver a second message regarding drug discovery and therapeutics. Although drug discovery efforts are generally geared towards targeting essential proteins, the effectiveness of this approach may be limited. Supplemental drugs that target non-essential proteins (e.g. Mfd) during the treatment of infections (or various diseases such as cancer) have the potential to significantly improve the efficiency and/or potency of current treatment regimens. Therefore, development of novel therapeutics targeting non-essential proteins could expand the arsenal of drugs available to combat AMR and potentially other diseases.

2.5 Figures

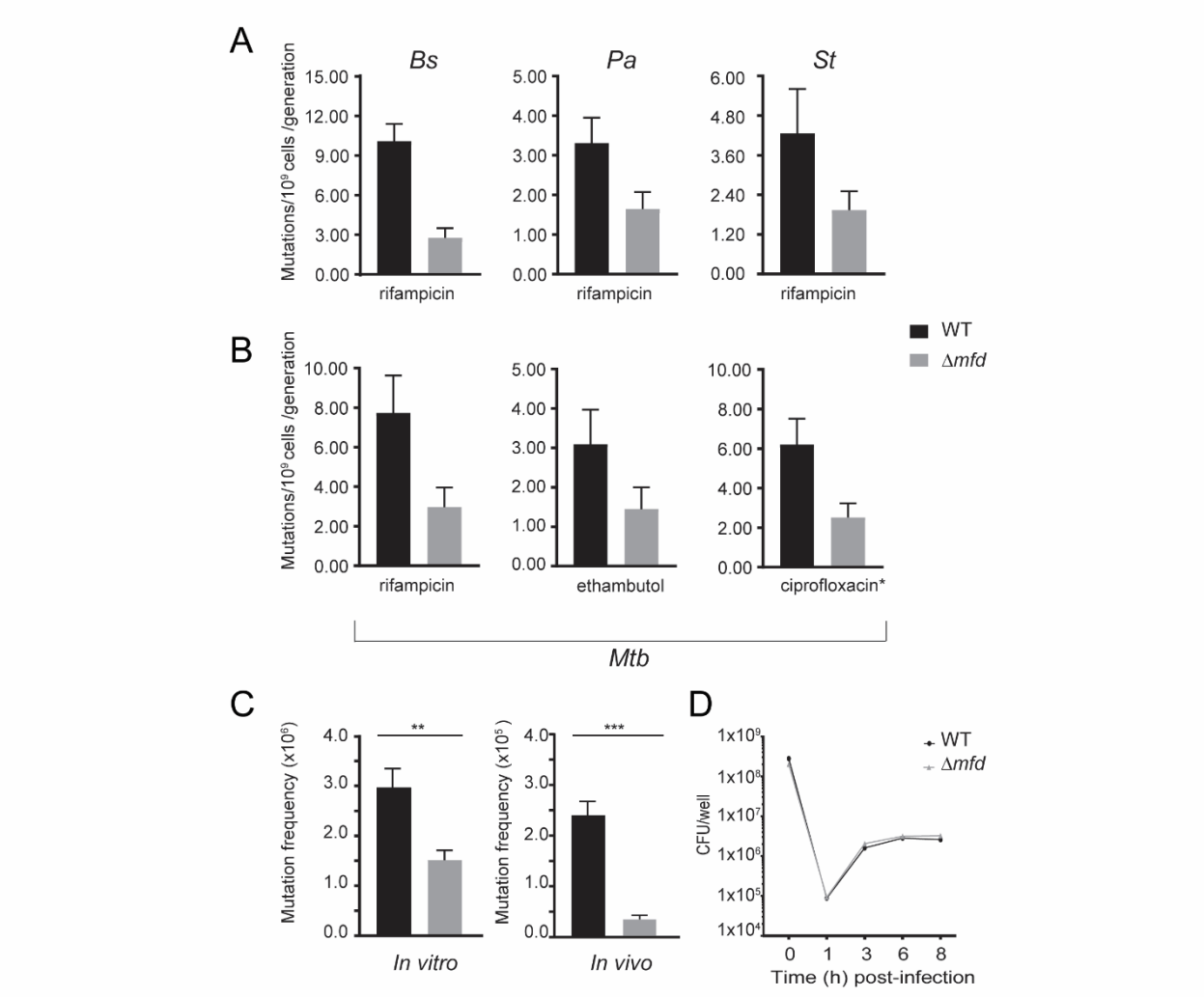

Figure 2.1 Mfd promotes mutagenesis in diverse bacterial species
(A) Mutation rates of WT (black) and Δ*mfd* (gray) strains to rifampicin for three indicated species (*Bs = B. subtilis* HM1, *Pa = P. aeruginosa* CF127, *St = S. typhimurium* ST19). Number of replicates for *Bs*=75, *Pa*=42, *St*=36. Error bars are 95% confidence intervals. (B) Mutation rates of *Mtb* (H37Rv) to three different antibiotics for WT (black) and Δ*mfd* (gray). Number of replicates for *Mtb*=33-48. Error bars are 95% confidence intervals. *Ciprofloxacin y-axis is mutations/10^8 cells/generation. (C) Mutation frequency of *S. typhimurium* in culture tubes and during infection of CACO-2 cells. Frequency was measured by plating on M9 glycerol with 5-flourocytosine for CFU enumeration. Error bars are standard error of the mean. Two-tailed student's T-test determined statistical significance (**p-value <0.01, ***p-value <0.001). (D) CFU enumeration of WT and Δ*mfd S. typhimurium* strains upon infection of CACO-2 cells.

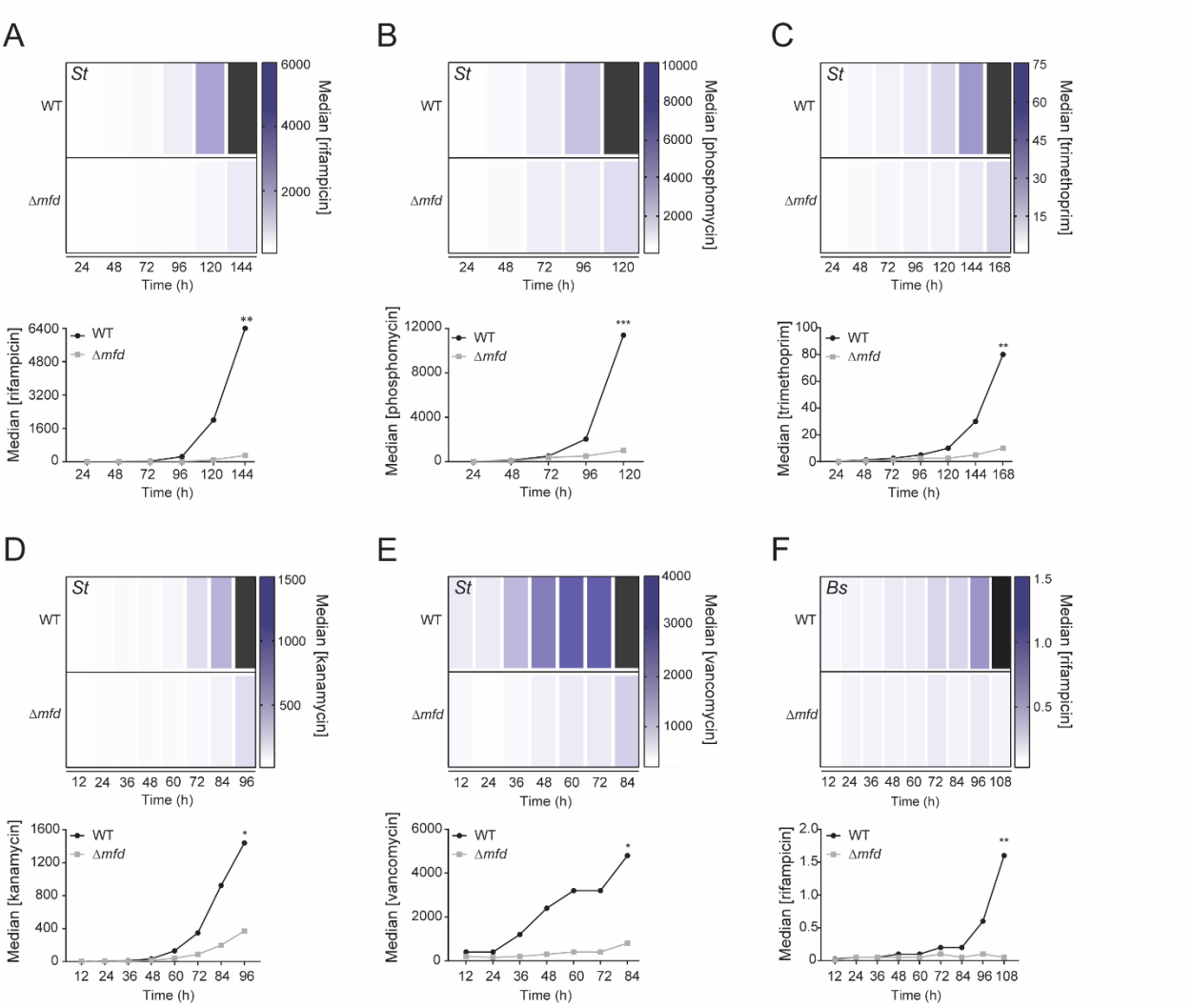

Figure 2.2 Mfd promotes evolution to various classes of antibiotics

Evolution of *S. typhimurium* ST19 (A-E) and *B. subtilis* HM1 (F) to indicated antibiotics. Heat maps and line plots show median antibiotic concentration for WT and *Δmfd* strains at each sampled time point. Black bars represent median growth greater than highest concentration shown on the scale. Concentrations for all antibiotics are in µg/mL. Statistical significance was determined using a two-tailed Mann–Whitney *U* test (*p-value <0.05, **p-value <0.01, ***p-value <0.001). Number of replicates for each strain and antibiotic of *St* and *Bs* are 12-30.

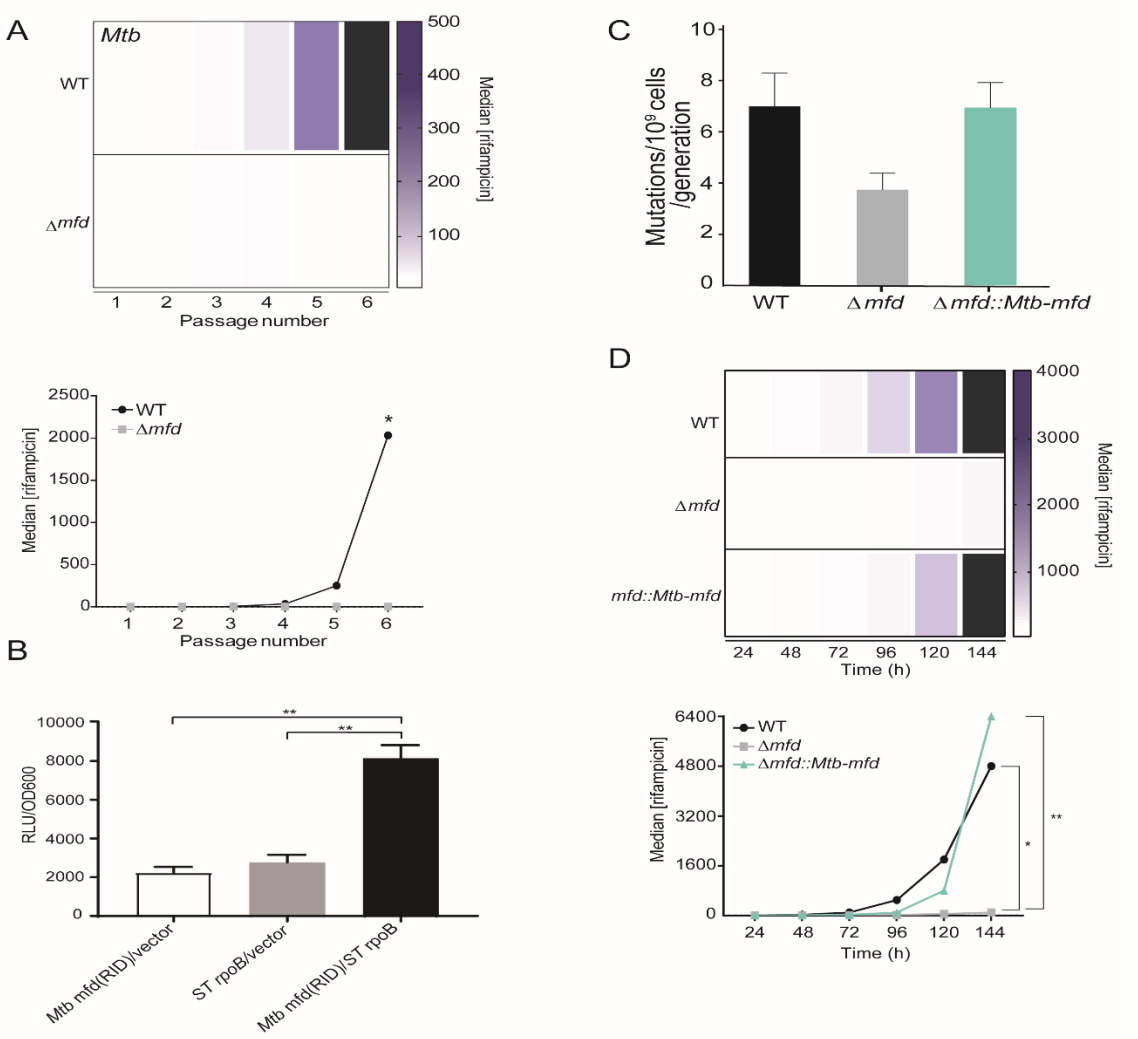

Figure 2.3 Mfd promotes evolution to antibiotics in *Mtb*

(A) Evolution of *Mtb* H37Rv to rifampicin. Heat maps and line plots showing median rifampicin concentration for WT and Δ*mfd* strains at each sampled passage from a representative experiment is shown. Black bars represent median growth greater than highest concentration shown on the scale. Concentrations are in ng/mL. Statistical significance was determined using a two-tailed Mann–Whitney *U* test (*p-value <0.10). Number of replicates for each strain of *Mtb* is 6. (B) *Mtb* Mfd and ST19 RpoB interact. Relevant domains of *Mtb* Mfd and *S. typhimurium* ST19 RpoB proteins were cloned into a luciferase based bacterial 2-hybrid system. Interactions between these respective protein domains were measured by luminescence and normalized to OD_{600}. Results are from three independent experiments and error bars indicate standard error of the mean. Statistical significance was determined using two-tailed Student's T-test (***p-value < 0.001). (C) Mutation rate analysis were performed with indicated strains of *S. typhimurium* to rifampicin as in Figure 2.1. Number of replicates is 36-96. (D) Evolution of indicated strains of *S. typhimurium* to rifampicin. Plots and statistical testing for evolution assays were performed as described in Figure 2.2. Number of replicates per strain is 12-24. *p-value <0.05 between WT and Δ*mfd* strains and **p-value <0.01 between Δ*mfd::Mtb-mfd* and Δ*mfd* strains.

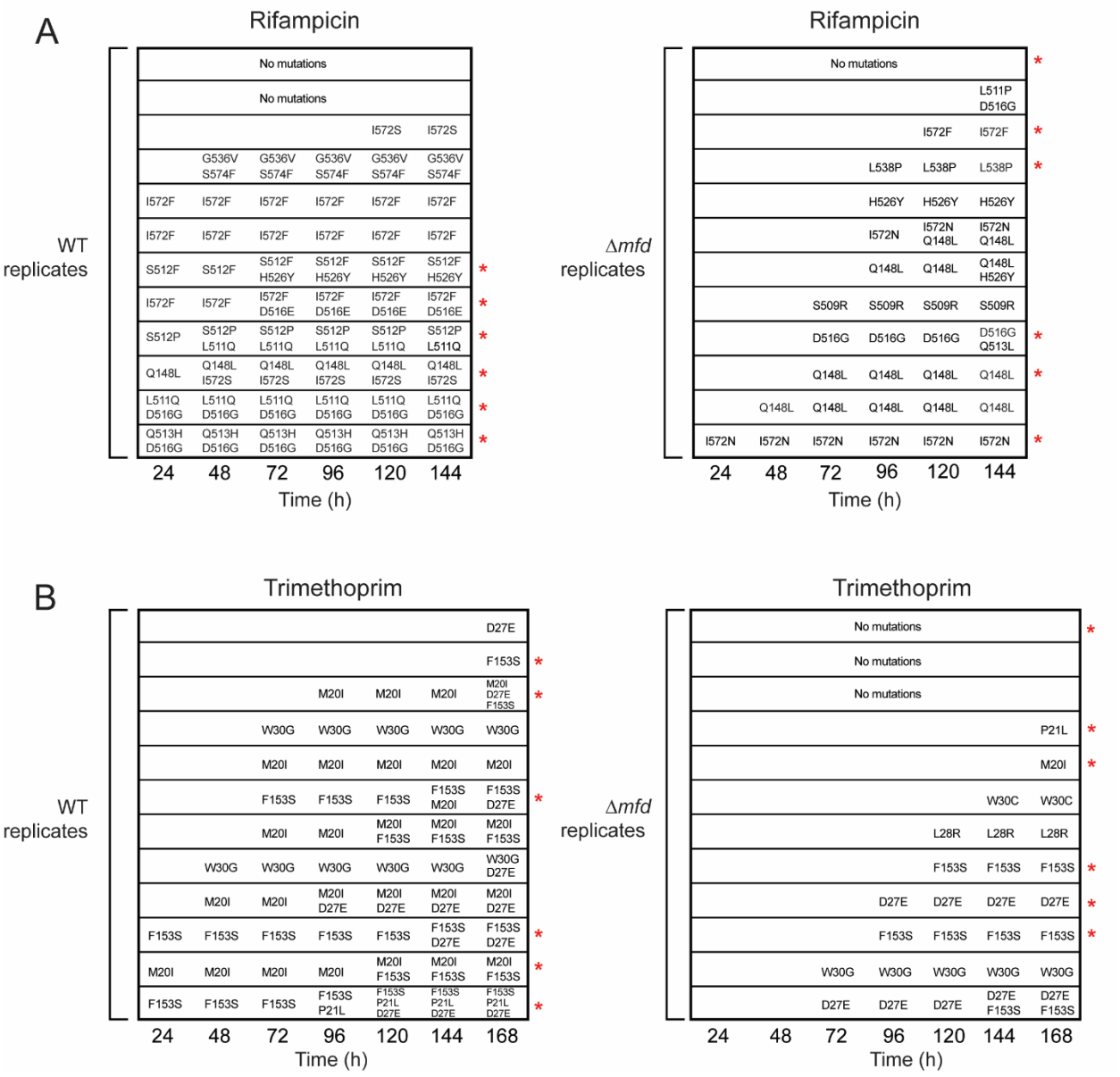

Figure 2.4 Cells lacking Mfd show fewer and delayed resistance-conferring mutations
(A) Sequencing of *rpoB* was performed at each time point from rifampicin evolution assays to identify mutations that confer resistance in WT and Δ*mfd* strains of *S. typhimurium* (12 replicates per strain). (B) Sequencing of *folA* was performed at each time point from trimethoprim evolution assays to identify mutations that confer resistance in WT and Δ*mfd* strains of *S. typhimurium* (12 replicates per strain). Shown are the position and corresponding amino acid changes. *Indicates replicates sequenced by whole genome sequencing.

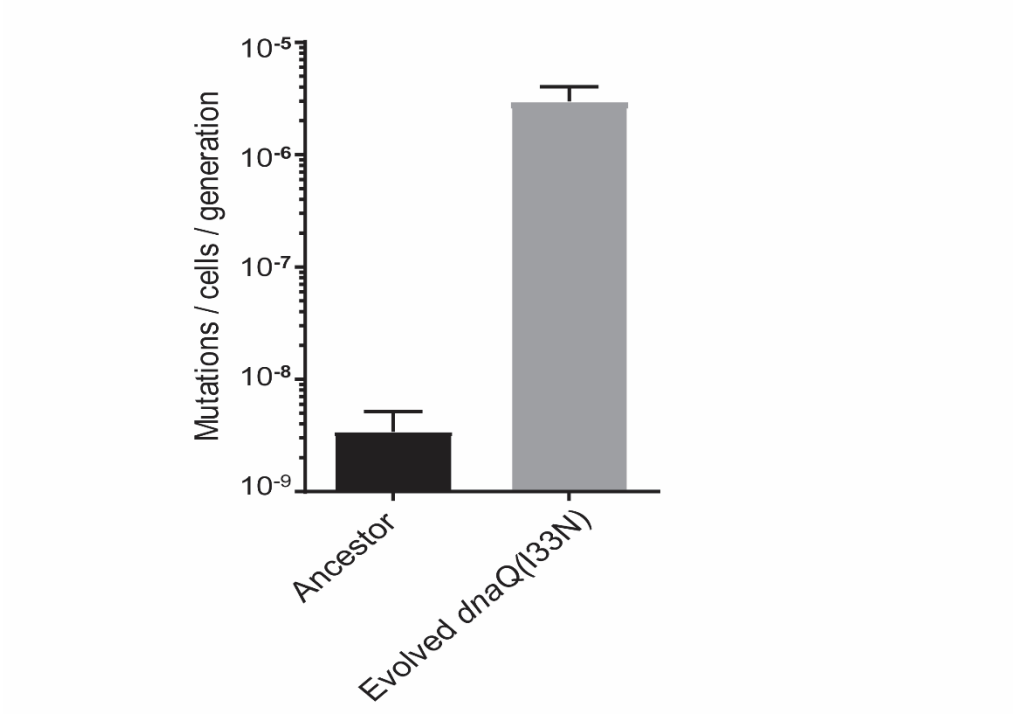

Figure 2.5 Development of hypermutation in evolved WT strains of *S. typhimurium*
Mutation rate analysis of *S. typhimurium* strains evolved to trimethoprim. Assays were performed on rifampicin plates as described in Figure 2.1. The number of replicates per isolate is 12.

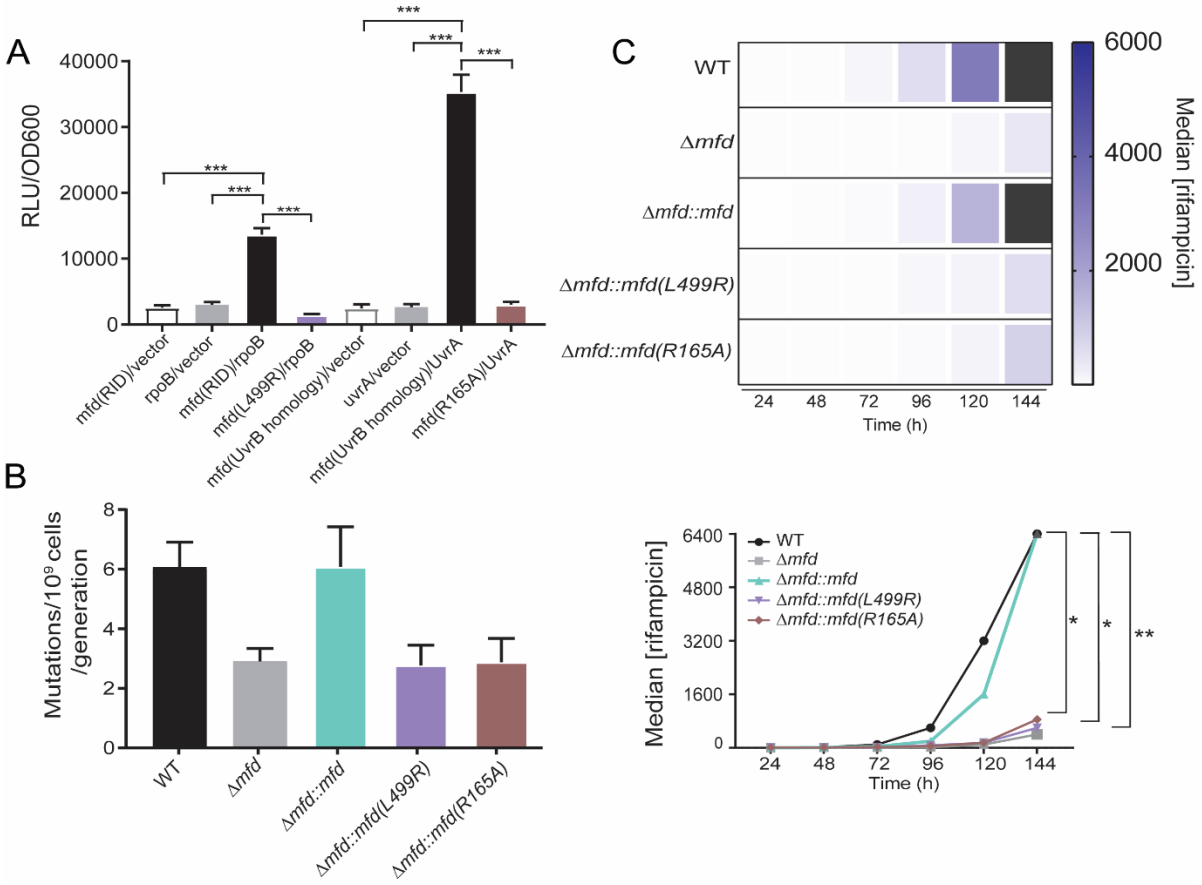

Figure 2.6 Mfd-RpoB and Mfd-UvrA interactions are essential for Mfd-driven mutagenesis and evolution to antibiotics

(A) Mutation of Mfd L499R and R165A residues abrogates RNAP and UvrA interactions, respectively. Relevant domains of the RpoB, Mfd, and UvrA proteins of *S. typhimurium* ST19 were cloned into a luciferase based bacterial 2-hybrid system. Interactions between the respective protein domains were measured as in Figure 2.3. Results are from three independent experiments and error bars indicate standard error of the mean. Statistical significance was determined using two-tailed Student's T-test (***p-value < 0.001). (B) Mutation rate analysis of indicated strains of *S. typhimurium* to rifampicin. Complement and point mutant (L499R and R165A) strains were expressed episomally. WT and Δ*mfd* strains contain pUC19 empty vector control. Number of replicates per strain is 36-112. Errors bars are 95% CI. (C) Evolution of indicated *S. typhimurium* strains to rifampicin. Complement and point mutant strains (L499R and R165A) were expressed episomally. WT and Δ*mfd* strains contain pMMB67EH empty vector controls. Strains were grown in 50 μg/mL carbenicillin to maintain selection of episomes. Plots and statistical testing for evolution assays were performed as described in Figure 2.2. Number of replicates per strain is 12-24. **p-value <0.01 between WT and Δ*mfd* strains and *p-value <0.05 between WT and Δ*mfd::*Δ*mfd*(L499R) and WT and Δ*mfd::*Δ*mfd*(R165A) strains.

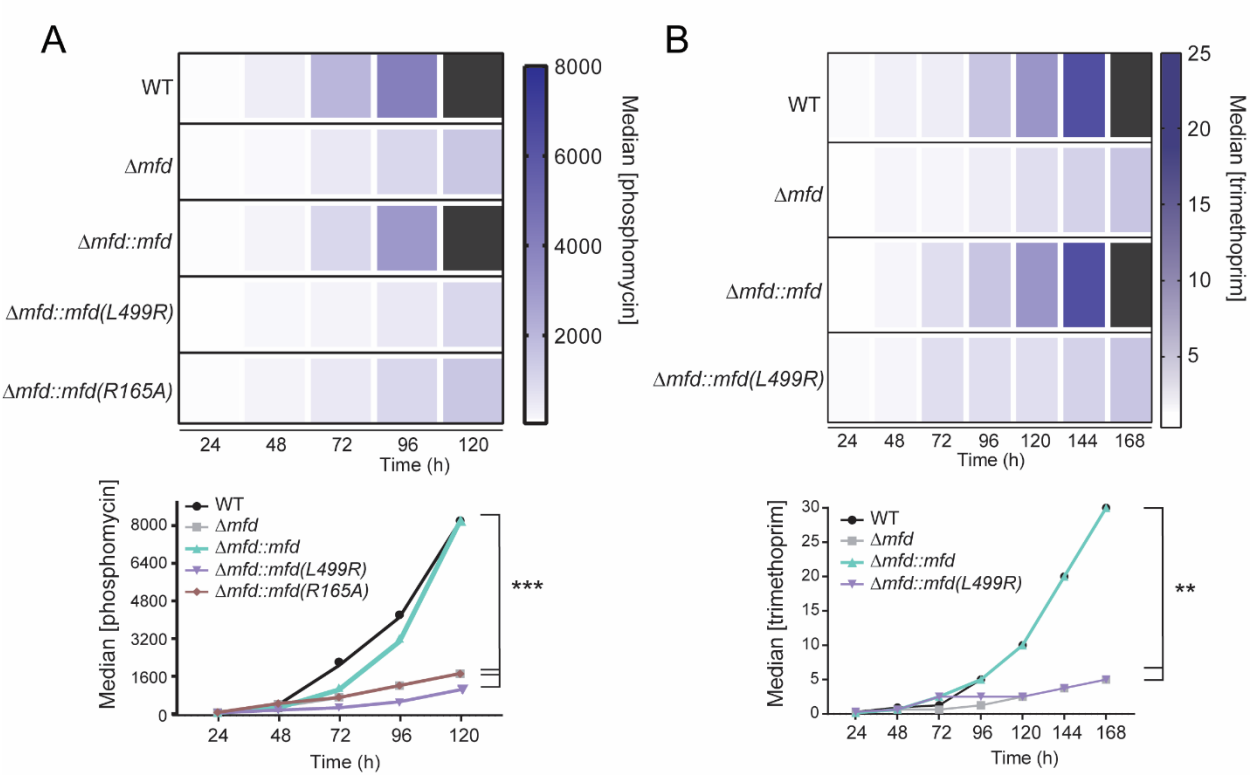

Figure 2.7 Mfd requires interaction with RNAP and UvrA to promote evolution to antibiotics Evolution of indicated *S. typhimurium* ST19 strains to phosphomycin (A) and trimethoprim (B). Plots and statistical testing for evolution assays were performed as described in Figure 2.2. ***p-value <0.001 between WT and Δ*mfd*, between WT and Δ*mfd::mfd*(L499R), and between WT and Δ*mfd::mfd*(R165A) strains for evolution to phosphomycin. **p-value <0.01 between WT and Δ*mfd* and between WT and Δ*mfd::mfd*(L499R) strains for evolution to trimethoprim. n = 12-24 replicates per strain.

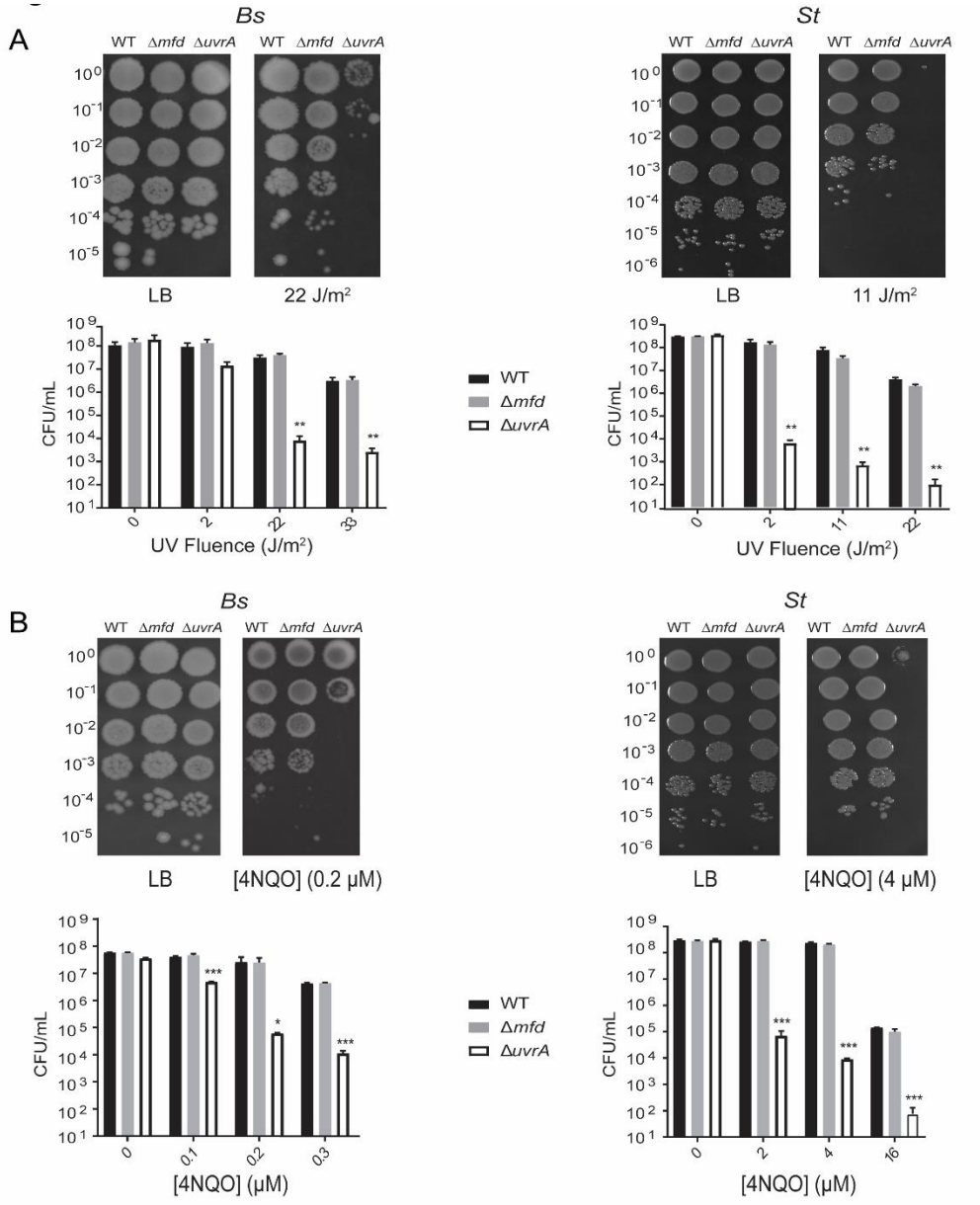

Figure 2.8 Cells lacking Mfd are not significantly sensitive to DNA damaging agents
Survival assays to (A) UV damage and (B) 4NQO for WT, Δ*mfd*, and Δ*uvrA* strains of *B. subtilis* HM1 (*Bs*) and *S. typhimurium* ST19 (*St*). (*uvrA* knockouts, known to be sensitive to DNA damage, were included for comparison to *mfd* knockouts). Data represents at least two independent experiments with duplicates for each experiment. Errors bars indicate s.e.m. Statistical significance was determined using two-tailed Student's t-test (*p-value <0.05, **p-value <0.01, ***p-value <0.001).

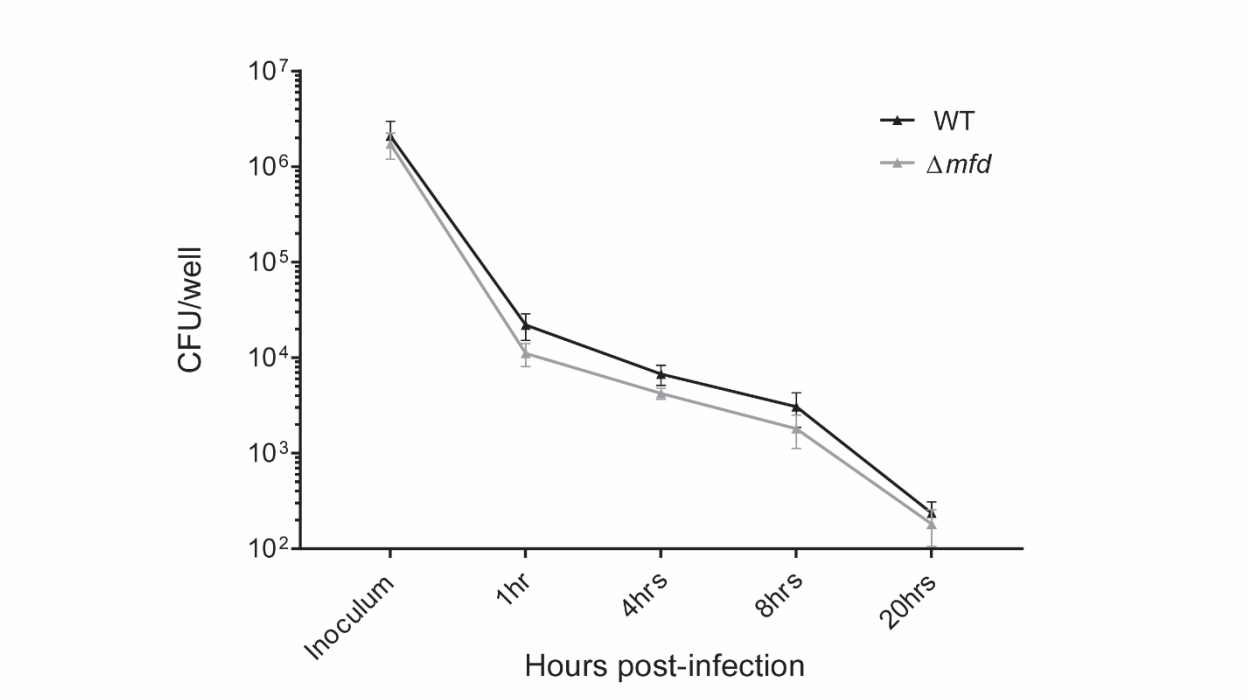

Figure 2.9 Strains lacking Mfd show no survival defects in bone marrow macrophages, Murine-derived bone marrow macrophages (BMMs) were infected with WT and Δ*mfd* strains of *S. typhimurium* ST19 and harvested for CFU enumeration at indicated times points. Data represents two independent experiments with triplicate samples for each given experiment. Error bars indicate s.e.m.

Chapter 3

Mfd is a critical RNAP co-factor that links structured RNAs to evolution

3.1 Summary

Regulatory RNAs are a ubiquitous and diverse class of molecules that regulate a myriad of biological processes, ranging from metabolism, to tRNA regulation, to toxin-antitoxin functions in bacteria. Regulatory RNAs often contain a degree of secondary structures, which can strongly impact RNA polymerase (RNAP) progression due to pausing and termination at such sites. How RNAPs proceed through these difficult-to-transcribe regions is not fully understood *in vivo*. Furthermore, a connection between RNA secondary structures and evolution has not been reported. Here were find that the evolvability factor and translocase, Mfd, regulates RNAP association with, specifically, regions containing highly structured regulatory RNAs, in both Gram-positive and Gram-negative bacteria. We identify sites in the genomes that are bound by Mfd, and where Mfd promotes RNAP release, and find that nearly all such sites contain regulatory RNAs, many of which contain predicted strong secondary structures. Our experiments strongly suggest that Mfd functions to terminate transcription at these sites, serving as a co-factor to RNAP at chromosomal regions expressing highly structured, regulatory RNAs. Furthermore, our data suggest that Mfd's regulatory function at these regions, including some toxin-antitoxin and metabolic genes, is essential. Lastly, our results indicate that Mfd promotes mutagenesis at regions containing structured RNAs. Based on these data, we propose that Mfd often serves as an essential RNAP co-factor at regions expressing highly structured RNAs, directly linking RNA secondary structure to evolution.

3.2 Introduction

Timely and efficient transcription is a fundamental requirement for maintaining cellular homeostasis. The process of transcription elongation is discontinuous, with RNA polymerase (RNAP) processivity altered by a wide range of obstacles. These obstacles range in severity, from pause sites that slow the rate of RNAP[157–159] to more severe obstacles, such as protein roadblocks and the replication fork, which induce the reverse translocation of RNAP with respect to both DNA and nascent RNA (termed RNAP backtracking)[160–163].

RNAP pausing can be promoted by the formation of stable RNA hairpin structures that form in the exit channel of RNAP and inhibit its movement[159,164–166]. A pause site may induce RNAP backtracking but perhaps more commonly induces a "half translocated" state of RNAP[165], which inhibits its immediate processibity[167]. Pausing via RNA secondary structure can regulate gene expression at riboswitches, can promote coupling of transcription and translation[168,169], and is critical for the process of intrinsic transcription termination in bacteria[121,170]. RNAP pausing is regulated directly by NusA and NusG factors, which oppose each other's functions, with RNAP pausing promoted by NusA and RNAP elongation promoted by NusG[171,172].

When RNAP encounters more severe roadblocks, such as a protein complexes, the enzyme can backtrack a significant distance on DNA. Backtracking of RNAP is both prevented and resolved through various mechanisms, including cellular factors that help re-establish RNAP elongation (i.e. antibacktracking factors)[118]. One of these factors is the translocase protein Mfd. *In vitro* studies show that Mfd utilizes its RNAP binding properties and forward translocase activity to rescue arrested RNAP and in doing so can restore RNAP to active elongation[110] as well as

promote RNAP termination[121]. While the biochemical characteristics of Mfd are relatively well defined, the context in which its translocase and antibacktracking capacity become important in cells remains elusive.

In vivo, Mfd was initially ascribed the function of a DNA repair factor that promotes transcription-coupled repair (TCR)[92,94,173,174]. In the TCR pathway, Mfd removes stalled RNAP at bulky DNA lesions to expose the offending lesion to the nucleotide excision repair (NER) pathway via its UvrA binding capacity. However, cells lacking Mfd show little to no sensitivity to DNA damaging agents that promote RNAP stalling[45,106], especially relative to other TCR factors[104,105]. Outside of its role in TCR, Mfd's biochemical activities are also thought to be important *in vivo* for preventing genomic instability induced by collision between the replication and transcription machineries[118] and for catabolite repression in *Bacillus subtilis*[108,175]. Interestingly, Mfd is also an evolvability factor in diverse bacterial species, promoting rapid evolution of antibiotic resistance development[45] as well as stationary-phase mutagenesis[43,87]. The mechanisms of Mfd-mediated mutagenesis and what sites in the genome are particularly prone to this mutagenesis, remains unclear.

Mfd is highly conserved, both functionally and structurally[45,176,177] and in many organisms its expression is constitutive. This implies a fundamental role for Mfd, and indeed, recent work suggests that the protein plays a housekeeping role in cells by associating with RNAP in the absence of exogenous stressors[112]. Yet what endogenous sites Mfd may act upon in a proposed housekeeping role is unclear as a comprehensive characterization of its *in vivo* functions remains lacking.

In this work, I attempt to define the *in vivo* context in which Mfd becomes essential for transcription and cellular viability. I discover that Mfd acts to promote RNAP release and transcription termination at highly structured regulatory RNAs. I find that Mfd's activity is important for regulating transcription at various critical loci in *B. subtilis*, including toxin-antitoxin systems, and that misregulation in the absence of Mfd can be lethal. In addition, I find that Mfd promotes mutagenesis at toxin-antitoxin loci, suggesting that Mfd may promote the evolution of critical regulatory RNAs. This work suggests that RNA secondary structure is a major impediment to transcription *in vivo* and that Mfd acts at these sites. In doing so, we propose that Mfd is an essential factor for both the transcriptional regulation and the evolution of critical regulatory RNAs.

3.3 Results

Mfd functions as an RNAP co-factor

In this work, we sought to better characterize the cellular function of the enigmatic Mfd protein. We began by assessing Mfd's genome-wide association using chromatin immunoprecipitation sequencing (ChIP-seq). We constructed a *B. subtilis* C-terminally Myc-tagged strain (Figure 3.2A) and exponentially growing cultures were subsequently harvested for ChIP-seq analysis. Under these conditions, we find pervasive Mfd association in the genome, compared to a ChIP-seq performed using Myc antibody in *B. subtilis* lacking a Myc-tagged Mfd (Figure 3.1A). Given these data and the known physical interaction between Mfd and RNAP[97], we hypothesized that Mfd functions largely as a co-factor of RNAP in cells, associating genome-wide through the process of transcription.

To test this hypothesis, we sought to determine whether Mfd binding peaks were correlated with RNAP association on the genome. We therefore performed ChIP-seq of RpoB, the β subunit of RNAP, in *B. subtilis*. We indeed find that Mfd association genome-wide is largely coupled to RpoB occupancy (Pearson coefficient 0.68) (Figure 3.1B, Figure 3.3), suggesting that Mfd may function as an RNAP co-factor in *B. subtilis*.

Mfd is highly conserved across all bacterial phyla, and functional homologs exist throughout all domains of life[103,113,178]. Experiments show that Mfd function is also highly conserved, at least with regards to its mutagenic activity[45]. To test whether Mfd's genome-wide coupling with RNAP was also conserved, we constructed an Mfd-Ypet fusion in the gram-negative pathogen *S. typhimurium* (Figure 3.2B) and performed Mfd and RpoB ChIP-seq experiments. As in *B. subtilis*, we find that Mfd and RpoB occupancy are highly correlated (Pearson coefficient 0.85) in *S. typhimurium* (Figure 3.4).

Mfd's binding to RNAP requires the RNAP interacting (RID) domain, and mutations in this domain abrogate this interaction[97]. We therefore hypothesized that disrupting the interaction between Mfd and RNAP would inhibit Mfd's association with DNA. We constructed a mutation in our Mfd-myc tagged strain (MfdL522R), which corresponds with the same amino acid substitution shown to disrupt Mfd binding in *E.coli* (L499R)[97]. We find that the MfdL522R mutant exhibits little to no ChIP-seq binding (Figure 3.1C), suggesting that interactions with RNAP are essential for Mfd association with DNA. These findings further suggest that Mfd functions a genome-wide RNAP co-factor *in vivo*.

Mfd requires transcription elongation for association with DNA

In vitro, Mfd helps promotes the rescue of arrested transcription elongation complexes (TECs), yet how Mfd recognizes stalled RNAP *in vivo* remains unclear. We therefore sought to discriminate whether Mfd association with DNA was facilitated via loading during transcription initiation or during the elongation phase of transcription. To distinguish between transcription initiation and elongation, we utilized the antimicrobial rifampicin, which directly blocks transcription initiation[179,180], subsequently eliminating the formation of transcription elongation complexes (TECs). Rifampicin treatment largely eliminated Mfd ChIP-seq binding signal (Figure 3.1D), showing that active transcription elongation is a requirement for Mfd association across the genome.

Mfd directly promotes release of RNAP

The co-factor activity of Mfd at elongating RNAPs *in vivo* suggests that Mfd may facilitate the activity of RNAP. Mfd can function to promote both transcription elongation as well as transcription termination, at least under *in vitro* conditions[110,121]. However, the importance of these functions of RNAP *in vivo* remains elusive.

We wondered if Mfd's close association with RNAP *in vivo* was relevant for transcription elongation, transcription termination, or both. We therefore performed ChIP-seq of RpoB in WT and *Δmfd* strains in *B. subtilis* to look for genes where RpoB occupancy was altered in the absence of Mfd. ChIP-seq reveals that RpoB occupancy at most genes is not significantly altered by Mfd activity (Figure 3.5). However, we additionally find genes with both increased as well as decreased RpoB occupancy in *Δmfd* (Figure 3.5A, Figure 3.6). We specifically notice a bias

towards a greater number of genes that exhibited increased RpoB occupancy (relative to genes with decreased RpoB occupancy), in *Δmfd*. Further quantification of these results reveals that in *Δmfd*, a total 116 genes- grouped into 71 transcription units (TUs)- exhibit at least two-fold increased RpoB occupancy and 54 genes- grouped into 30 TUs- exhibit at least two-fold decreased RpoB occupancy (Tables 3.1, 3.2, and 3.3).

We next wanted to determine whether the changes in RpoB occupancy observed in *Δmfd* were directly due to Mfd's activity at those regions. When comparing genes associated with Mfd (defined as genes with an Mfd ChIP association one standard deviation greater than the mean) to RpoB association, we observe a greater Mfd association with genes that are highly expressed, as expected (Figure 3.5B). Critically, we also observe Mfd association many of the genes with increased RpoB occupancy in *Δmfd*, but not at genes with decreased RpoB occupancy in *Δmfd* (Figure 3.5B). Quantification shows that 52 of the 116 genes (and 35 of the 71 TUs) with increased RpoB occupancy in *Δmfd* are bound by Mfd, while none of the 53 genes with increased RpoB occupancy are bound by Mfd. From these findings we conclude that Mfd likely functions to promote the termination of elongating RNAP *in vivo*.

Regions where Mfd promotes decreased RNA occupancy are enriched for structured regulatory RNAs

In vitro, Mfd's translocase activity can help release RNAP exposed to various obstacles. However, the endogenous obstacles that Mfd helps RNAP overcome remains unknown. We wondered if release of RNAP by Mfd detected by our ChIP assays revealed endogenous sites of RNAP pausing or stalling. This model would explain the enhanced Mfd binding at these sites

(due to decreased RNAP processivity). It would also explain the decreased RpoB occupancy in the presence of Mfd, as Mfd could be promoting translocation and release of RNAP as a result of its decreased processivity.

Intriguingly, we noticed that of the 35 TUs that show both an increase in RNAP density in *Δmfd* and direct Mfd association, a striking 32 of those (92%), contained at least one regulatory RNAs (Table 3.4). These regulatory RNAs are a subset of the 1583 regulatory RNAs in *B. subtilis*, which encompass a wide variety of RNAs, including independent, non-coding transcripts, antisense RNAs, and multiple riboswitches[181]. We hypothesized that the mechanism underlying the increase in RNAP association at the identified TUs in *Δmfd* was related to the RNA secondary structure impeding RNAP processivity. Previous work characterized the predicted secondary structure for each regulatory RNA in *B. subtilis*[182], and we sought to test whether regions with increased RNAP in *Δmfd* were more prone to forming secondary structure. We compared the average minimum free energy (MFE) z-score, a proxy of RNA structure stability[183,184], between the 42 regulatory RNAs in TUs that show increased RpoB density in *Δmfd* and Mfd association to those that show no difference in RpoB density between WT and *Δmfd*. We find that regulatory RNAs associated with Mfd binding and increased RpoB density in *Δmfd* have significantly higher predicted RNA secondary structure (Figure 3.7). In comparison, only 12 of the 31 TUs with decreased RNAP density in *Δmfd* contain regulatory RNAs (Table 3.5), a number consistent with the average percentage of TUs expected to contain regulatory RNAs in *B. subtilis*. Critically, these regulatory RNAs do not exhibit a significant increase in predicted secondary structure (Figure 3.7).

RNAP termination at structured regulatory RNAs is specific to Mfd

Various factors are known to help rescue arrested RNAP through different mechanisms. One of the most well-known and highly conserved is the GreA protein, which functions as an RNAP antibacktracking by cleaving the nascent 3' RNA that has extruded from the RNAP catalytic channel during backtracking[118,185].To test whether the antibacktracking activity of GreA also contributed to RNAP release at structured RNAs, we performed RpoB ChIP-seq of *B. subtilis* WT and *ΔgreA* (Figure 3.8). We find that in contrast to cells lacking Mfd, cells lacking GreA have only 12 genes with greater than two-fold RNAP occupancy (data not shown), which can be grouped into six TUs. Two of the TUs contain regulatory RNAs, neither of which contain significant predicted secondary structure. Interestingly, we find a total 469 genes exhibiting less than two-fold RpoB occupancy in *ΔgreA* (data not shown). This finding suggests that GreA mainly functions *in vivo* to promote transcription elongation and is important for promoting RNAP release at sites of high RNA secondary structure.

Mfd promotes transcription repression at sites of structured sRNAs

We next tested the effect of Mfd on transcription at sites containing structured RNAs. We began by performing RNA-seq of WT and *Δmfd* strains in *B. subtilis*. As is the case with our WT and *Δmfd* RpoB ChIP-seq results, we find more genes upregulated than downregulated in *Δmfd* (240 genes upregulated compared to 138 genes downregulated) (Figure 3.9A and table not shown). In addition, of the 116 genes that have greater than 2-fold RpoB signal in *Δmfd,* we find that 30 of these directly show increased expression, while none show decreased expression. Because standard RNA sequencing protocols are often not suitable for accurate measurement of small RNAs[186], we wondered if perhaps there were other genes that showed increased RpoB density in

Δmfd but were not detected in our RNA-seq analysis. We therefore directly measured RNA levels at three loci containing non-coding RNAs (the *trnY* locus, *txpa-ratA*, and *bsrH-asBsrH*), all of which show Mfd binding and increased RpoB signal in *Δmfd* using qRT-PCR. We find of all three of these loci have increased gene expression in *Δmfd* (Figure 3.9B). Interestingly, our RNA-seq results also show a significant increase in the expression of various prophage and mobile elements in *B. subtilis*, suggesting that the absence of Mfd may induce cellular stress. These findings suggest that Mfd's *in vivo* activity in reducing RNAP leads directly to the repression of transcription.

Transcriptional regulation by Mfd at toxin-antoxin loci is essential for cell survival

We next wanted to better define the physiological importance of Mfd at sites of structured regulatory RNAs. We began by focusing on the highest structured regulatory RNAs which had altered RpoB density in the absence of Mfd and were directly bound by the protein. These were two pairs of type I toxin-antitoxin (TA) loci in *B. subtilis* the *txpA/ratA* locus and the bsrH/as-bsrH locus. Type I TA loci are characterized by the expression of a small toxic peptide and a noncoding RNA that neutralizes toxin expression by direct binding and either inhibiting translation or promoting degradation of the toxin mRNA[187]. The cellular functions of type I TA loci remain unclear but they have been proposed to be important for diverse aspects of physiology, including persister formation[188], biofilm formation[189], and prophage maintenance[190]. Five type I TA loci have been identified in *B. subtilis*[187] and we observed that Mfd binding and at least two-fold increased RpoB signal in *Δmfd* at three of these (one other TA locus, *yonT/as-YonT* locus, shows a ~1.7-fold increased in RpoB density in *Δmfd*) (data not shown).

The regulation of type I TA loci is essential for cell survival, as overexpression of type I toxins can be lethal[191]. We hypothesized that Mfd's regulation of transcription at TA loci was important for survival. We therefore overexpressed the TxpA toxin under an IPTG (isopropyl b-D-1-thiogalacto-pyranoside) inducible promoter in both WT and *Δmfd* strains and performed cellular viability assays. We find that cells lacking Mfd are highly sensitized to both chronic (Figure 3.10A) and acute (Figure 3.10B) overexpression of TxpA. Cells lacking Mfd show up to five orders of magnitude sensitivity to overexpression. We next overexpressed the BsrH toxin in an inducible fashion in WT and *Δmfd* to test whether Mfd's effect was conserved. Indeed, we see sensitization of *Δmfd* with both chronic (Figure 3.10C) and acute (Figure 3.10D) overexpression of BsrH, with up to four orders of magnitude sensitivity. To confirm that this effect was directly due to overexpression, we performed qRT-PCR analysis in our WT and *Δmfd* strains containing our overexpression constructs. We confirm that in both TxpA and BsrH strains, toxin overexpression is induced by ~3-fold in *Δmfd*, both under minimal induction (no IPTG) and full induction (1mM IPTG), (Figure 3.10E) suggesting that Mfd's sensitivity is directly due to toxin overexpression.

Mfd promotes mutagenesis at toxin-antitoxin loci

We next wondered in Mfd's mutagenic activity was relevant for mutagenesis at the TA loci tested. Mfd promotes mutagenesis under various conditions, including starvation, stationary-phase growth, and in the context of replication-transcription conflicts. Most of these studies utilize engineered reporter systems to study Mfd-mediated mutagenesis, and it remains unclear what endogenous sites are prone to Mfd's effects on mutagenesis. Mfd also promotes antimicrobial resistance development, and our data show that the endogenous loci that help

confer antibiotic resistance are bound by Mfd. Given our findings that Mfd is critical for transcriptional regulation at structured RNAs, we wondered if such sites were particularly prone to Mfd-mediated mutagenesis. We therefore utilized our TA overexpression constructs to test this hypothesis. We performed Luria-Delbrück fluctuation assays to assess the rate of reversion at our TxpA and BsrH reporters. We plated cells on a high concentration of IPTG (1mM) to ensure that only genetic revertants would produce viable colonies. We find that the rate of reversion of the TxpA and BsrH constructs are markedly lower in *Δmfd* (Figure 3.11). These differences are higher than previously reported mutation rate reductions in *Δmfd* strains[45,85]. We subsequently isolated and sequenced 8 colonies at random from both WT and *Δmfd* strains to map where the mutations were occurring. We sequenced both the native TA loci as well as the engineered overexpression locus for both TA systems, as well as the promoters at both sites. All mutations identified were mapped to the overexpression loci and included mutations in the coding sequence of the *txpA* locus (Table 3.6). We conclude from these findings that Mfd promotes mutagenesis at sites of high secondary structure.

3.4 Discussion

In this work, we unravel the fundamental importance of RNA secondary structure on the *in vivo* function of Mfd. At highly structured RNAs, Mfd can function both as a termination factor as well as an evolvability factor[45], simultaneously promoting transcription regulation and mutagenesis (Figure 3.12).

Given the wide range of biological functions inherent to structured RNAs, the consequences of Mfd's activity at these sites are likely to be broad ranging. For example, we identify multiple

riboswitches containing TUs which Mfd bind to and promotes RNAP termination. These TUs are involved in many critical metabolic processes, ranging from beta-glucoside metabolism (*bglP-bglH-yxiE*)[192,193], to the utilization of glycerol (*glpF-glpK* and *glpT-glpQ*)[194] to purine metabolism (*purEKBCSQLFMNHD*)[195,196]. We also identify Mfd binding and RNAP termination at tRNA loci (the *trnY* locus in *B. subtilis* containing a highly structured RNA of unknown function), and a cell wall binding protein thought to play a role cell wall homeostasis, which in turn is regulated by a cis-acting antisense RNA (*YabE*/S25)[197], amongst many other sites. The physiological relevance of Mfd's activity at many of these sites requires further investigation. Additionally, given our findings in multiple bacterial species and the high conservation of Mfd, the link between RNA secondary structure and Mfd mediated transcriptional regulation and mutagenesis is also likely to be conserved.

Our work also expands on the underlying, and still enigmatic, *in vivo* function of Mfd. It has been proposed that Mfd can promote transcription termination at significant protein roadblocks[93,108,111,118,175,198], and while our findings do not contradict these results, they suggest that inherent secondary structure may be a more relevant site of Mfd activity *in vivo*. Interestingly, RNAP pausing at secondary structures may not promote a severely backtracked TEC but rather a less severe pre-transclocation state[157,167]. Although our experiments do not allow us to identify the precise state of RNAP at sites of Mfd activity *in vivo*, given the current models regarding RNAP activity findings hint at the possibility that Mfd can recognize and act not only upon RNAP that is backtracked but also RNAP that is paused and not fully backtracked, or perhaps simply decreased in its elongation rate.

Various mechanisms of transcription-associated mutagenesis (TAM) exist[70,72]. Based on our findings, we propose that the inherent structure of RNA may be an additional novel mechanism by which transcription promotes mutagenesis, specifically through Mfd's mutagenic activity. Interestingly, RNA secondary structure has been reported to enhance mutation rates in replicating retrovirus[199], suggesting that evolution via secondary structure may be a universal mechanism. Moreover, a recent study by Thornlow, et al., using computational analyses, revealed that tRNAs have increased mutation rates relative to other parts of the genome[200]. They also suggest that this phenomenon is linked to transcription. It is therefore quite possible that, at least in bacteria, the evolution of tRNA structures is mediated by Mfd.

By promoting DNA mutagenesis at sites of highly structured RNAs, Mfd may inherently alter the secondary structure encoded at the site of its activity, leading to novel or altered functions of the RNA. Our proposed model suggests that the secondary structure of the RNA itself is the driving force for TAM. Intriguingly, this model implies that mutation which alter the stability of the RNA structure, by either increasing or decreasing it, may directly modulate the mutation rate and influence evolution.

Figures and Figure legends

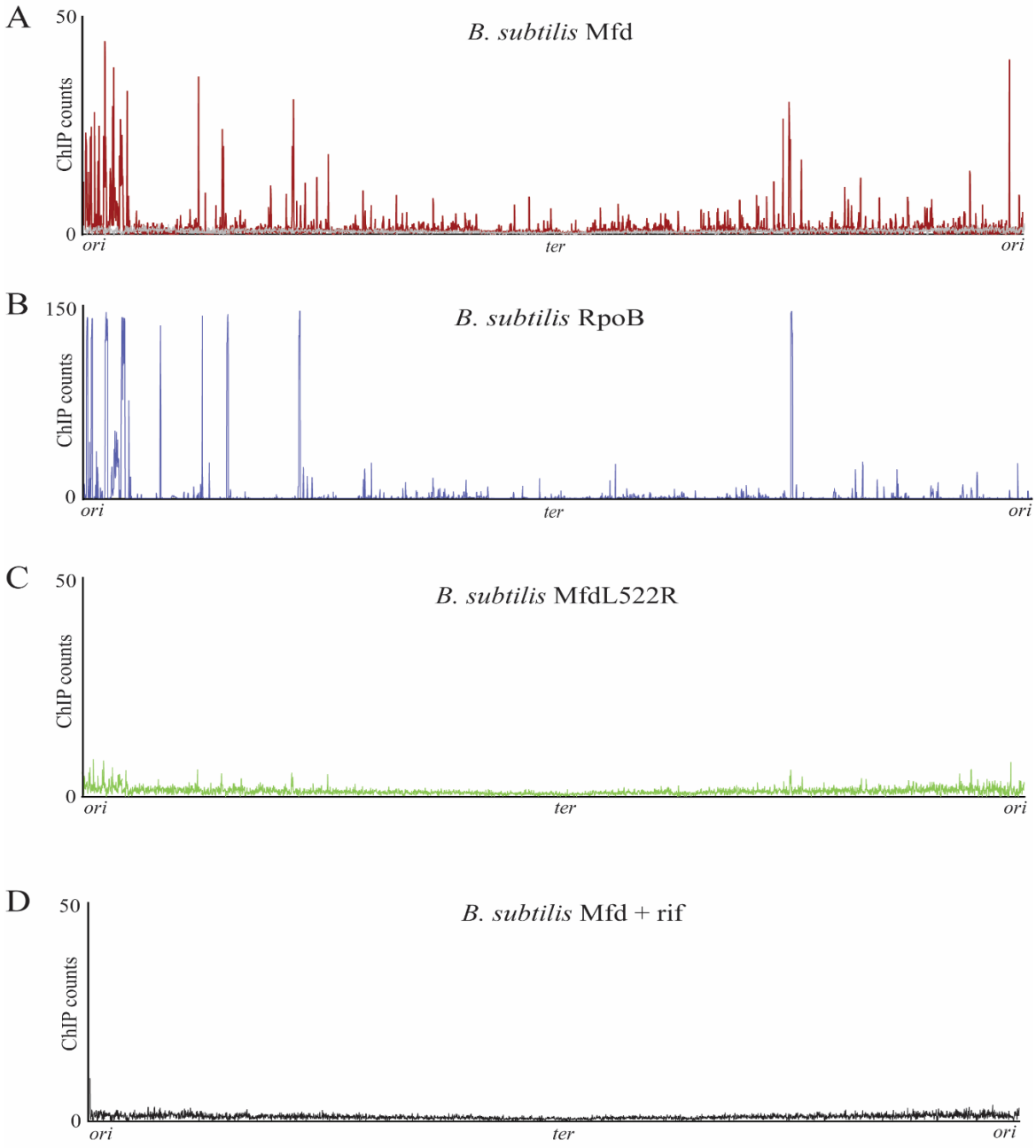

Figure 3.1 Mfd functions as an RNAP co-factor and requires transcription elongation for association with DNA

(A) ChIP-seq plot of *B. subtilis* Mfd-myc (read counts in red) and WT *B. subtilis* lacking an Mfd-myc tag (read counts in grey). (B) ChIP-seq plot of WT *B. subtilis* RpoB. (C and D) ChIP-seq plot of *B. subtilis* MfdL522R-myc point mutant and *B. subtilis* Mfd-myc treated with 50µg/mL of rifampicin for five minutes. Plots are normalized to total DNA input controls and are the average of at least two independent experiments

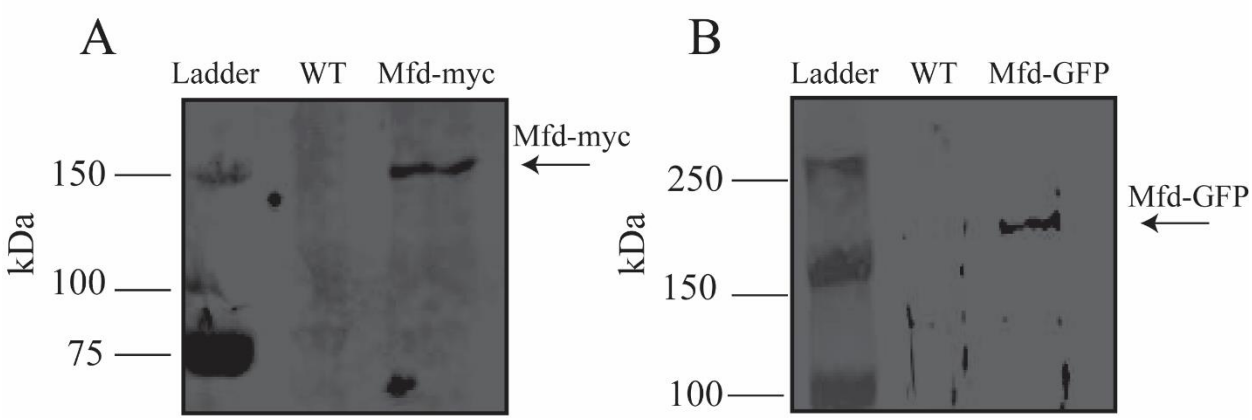

Figure 3.2 Western blots of *B. subtilis* Mfd-myc and *S. typhimurium* Mfd-Ypet
Western blot of (A) *B. subtilis* WT and Mfd-myc and (B) *S. typhimurium* WT and Mfd-Ypet. Anti-c-Myc antibody and anti-GFP antibody were used to probe blots (A) and (B), respectively.

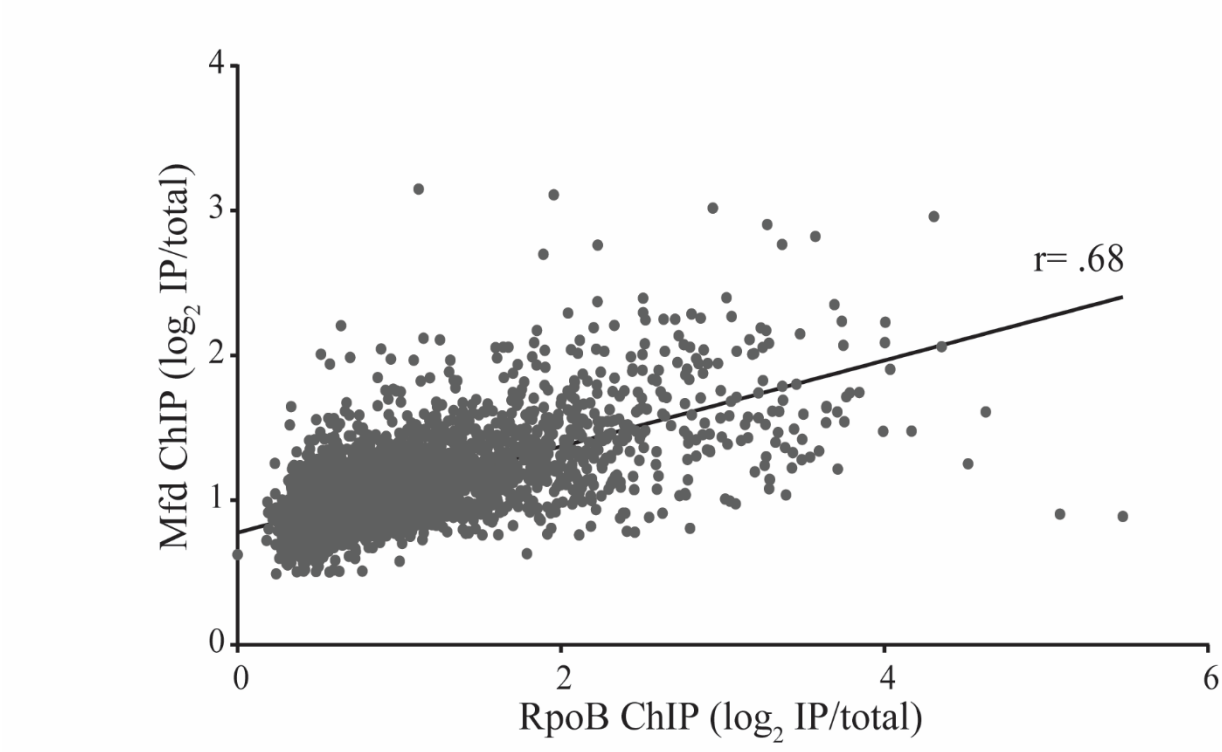

Figure 3.3 *B. subtilis* Mfd and RpoB ChIP-seq are correlated
Linear regression analysis comparing binding of Mfd and RpoB at each gene in *B. subtilis*. Mfd-myc ChIP-seq (from Mfd-myc tagged *B. subtilis*) and RpoB ChIP-seq (from WT *B. subtilis*) read counts were determined for each gene in *B. subtilis* and normalized as described in Figure 3.5. Pearson's correlation coefficient for *B. subtilis* Mfd and RpoB = 0.68.

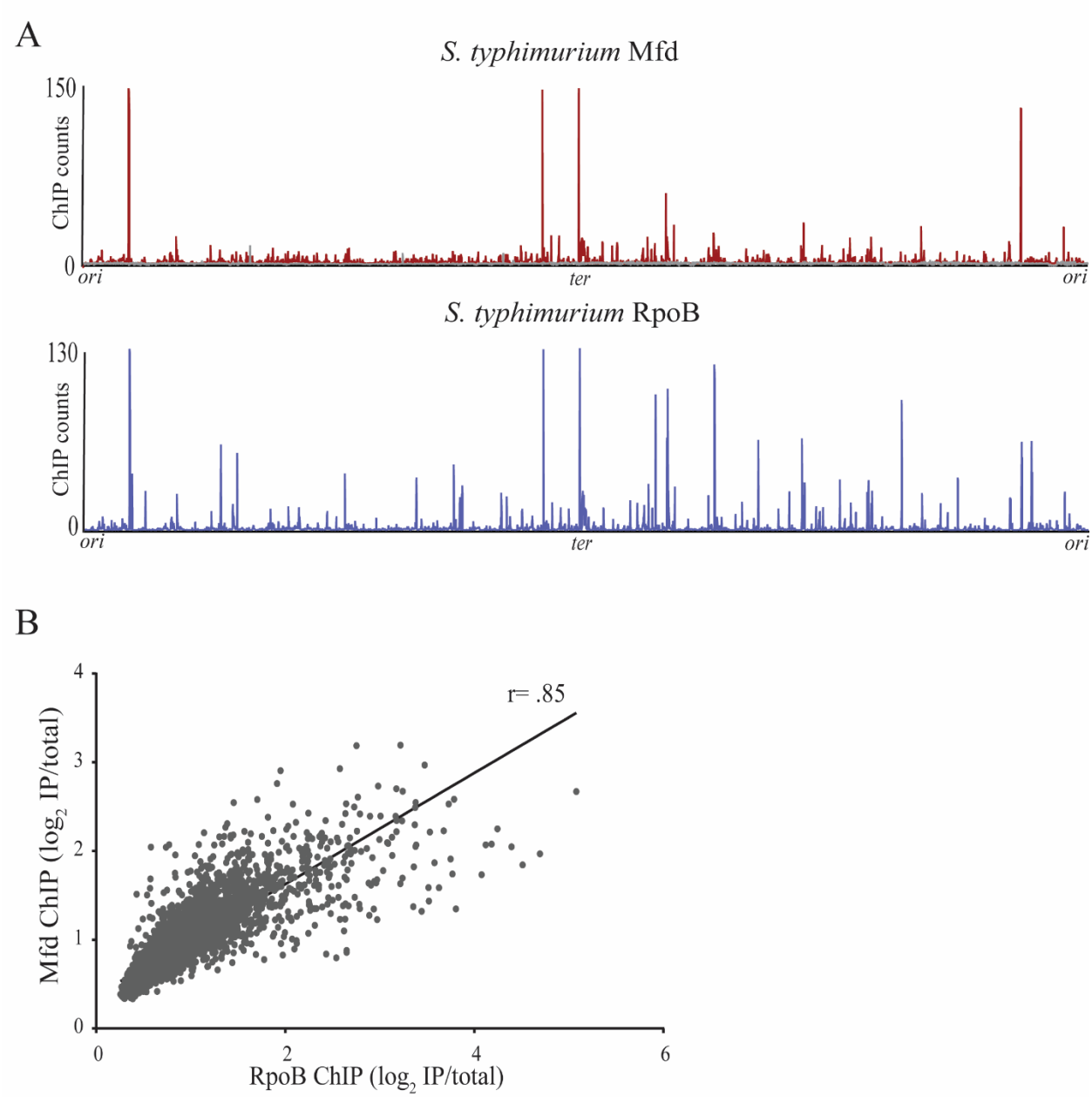

Figure 3.4 Mfd functions as an RNAP co-factor in *S. typhimurium*
(A) ChIP-seq plots of *S. typhimurium* Mfd-Ypet (read counts in red) and WT *S. typhimurium* lacking an Mfd-Ypet tag (read counts in grey). (B) ChIP-seq plot of *S. typhimurium* RpoB. (C) Linear regression analysis comparing binding of Mfd and RpoB at each gene in *S. typhimurium*. Read counts were determined and normalized as described in Figure 3.5. Pearson's correlation coefficient for *S. typhimurium* Mfd and RpoB = 0.85.

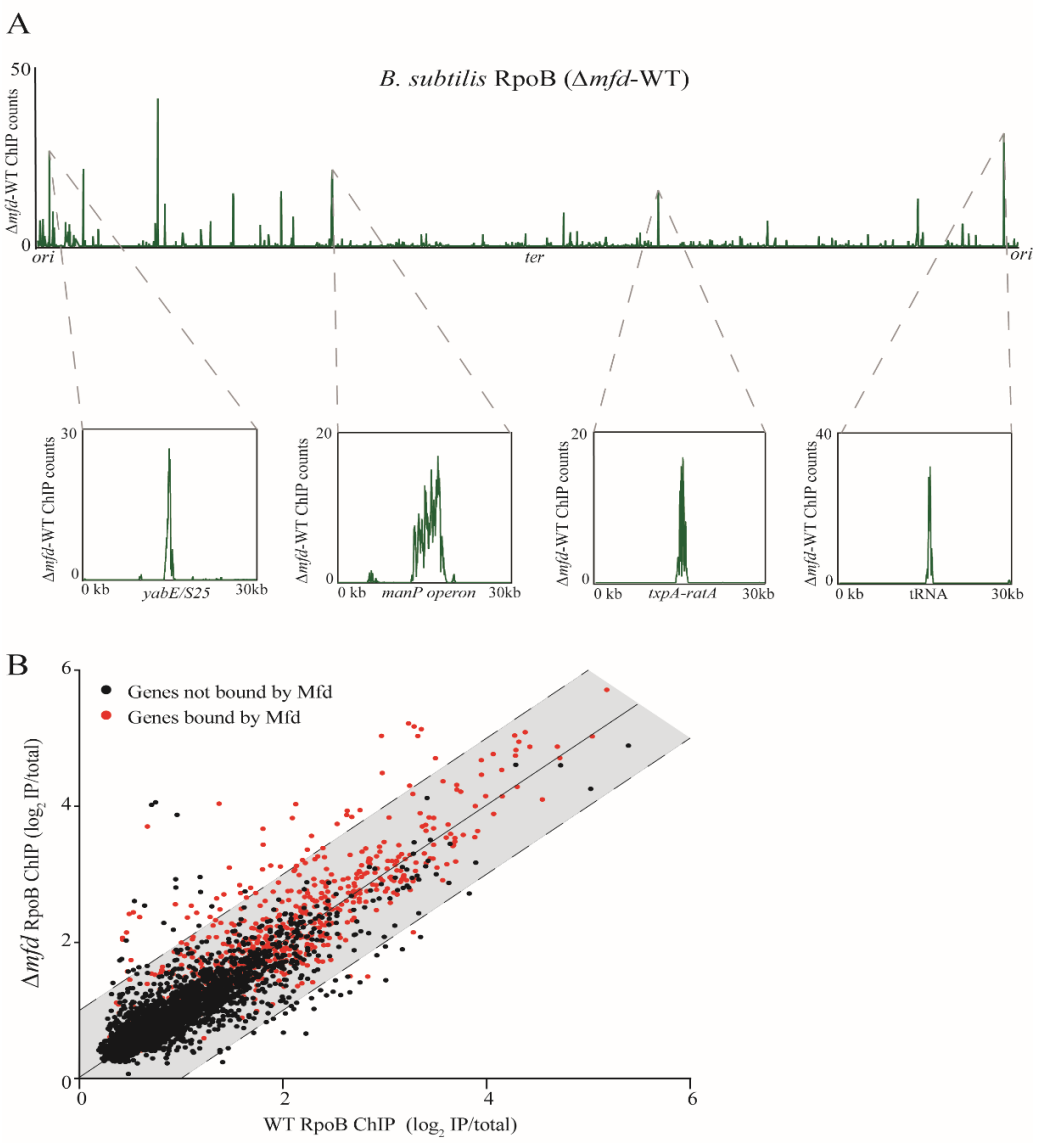

Figure 3.5 Mfd directly promotes release of RNAP *in vivo*
RpoB ChIP-seq plots showing regions of RpoB enrichment in *Δmfd*. WT and *Δmfd* RpoB ChIP-seq read counts were normalized to total DNA input and normalized read counts from WT were subtracted from *Δmfd*. High background signal from ribosomal RNA was removed from plots for better visualization. Zoomed in plots are representative regions of high RpoB enrichment in *Δmfd*. (B) Scatter plot of WT and *Δmfd* RpoB ChIP-seq measuring signal at each gene in *B. subtilis*. For quantification of ChIP signal at each gene, read counts for each gene were normalized to total library counts and IP samples were normalized to total DNA input to calculate an IP/Total DNA ratio. Ratios were $\log_2$ normalized and averaged over at least two independent experiments. Data points above and below colored shading indicate greater than two-fold increase and decrease in RpoB signal in the *Δmfd* strain, respectively. Data points in red indicate genes bound by Mfd. Binding is defined as one standard deviation greater than the average ChIP signal across all genes in *B. subtilis*. Calculation of Mfd binding at each gene was determined as described for RpoB ChIP samples.

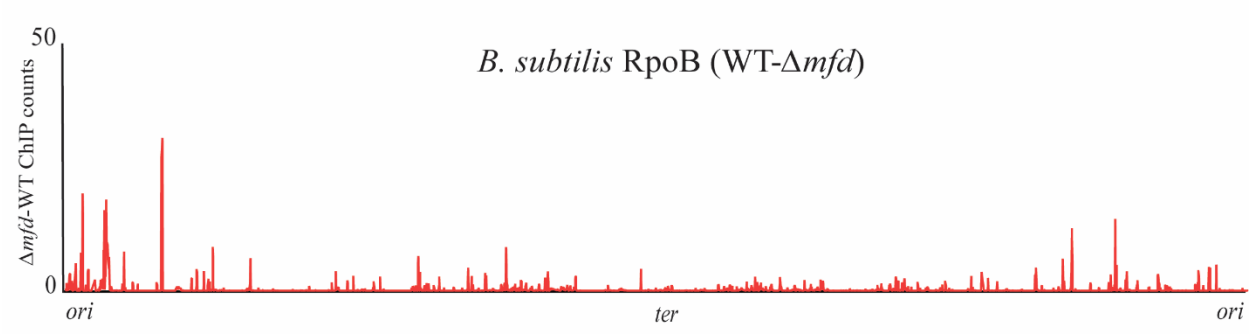

Figure 3.6 Sites of RpoB enrichment in WT B. subtilis compared to *Δmfd*.
RpoB ChIP-seq plots showing regions with increased RpoB signal in WT relative to *Δmfd* in *B. subtilis*. WT and *Δmfd* RpoB ChIP-seq read counts were normalized as in Figure 3.5 and read counts from *Δmfd* were subtracted from WT. High background signal from ribosomal RNA was removed from plots for better visualization.

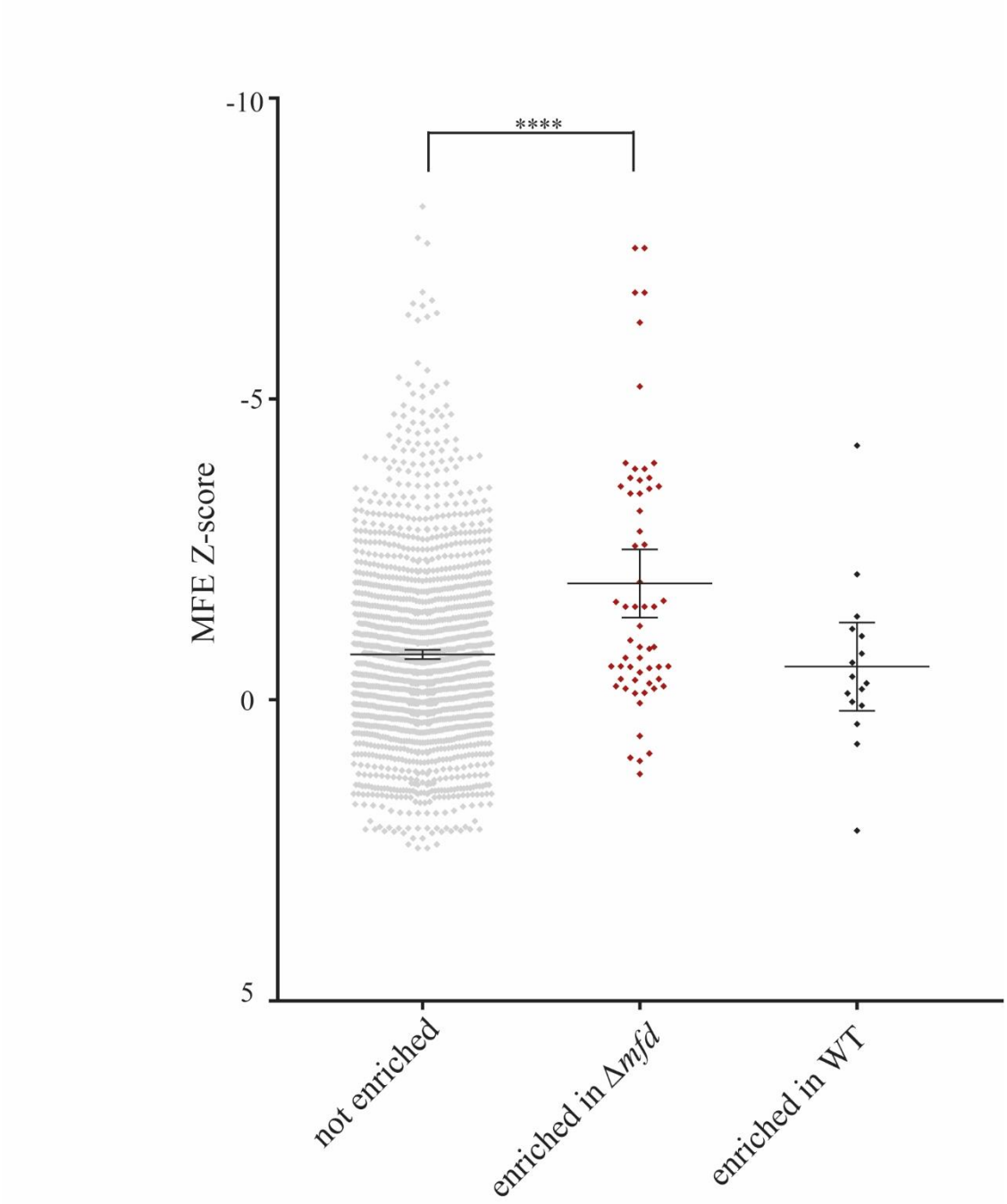

Figure 3.7 Transcription units with Mfd binding and increased RNAP density in *Δmfd* are enriched for structured regulatory RNAs

Scatter plot of the minimum free energy (MFE) Z-score for regulatory RNAs in *B. subtilis*. Data points represent regulatory RNAs within TUs that have no observed change in RpoB density between WT and *Δmfd* (grey data points), increased RpoB density in *Δmfd* and bound by Mfd (red data points), and decreased RpoB density in *Δmfd* (black data points). Statistical significance was determined using two-tailed Z-test for two population means ****P<0.0001.

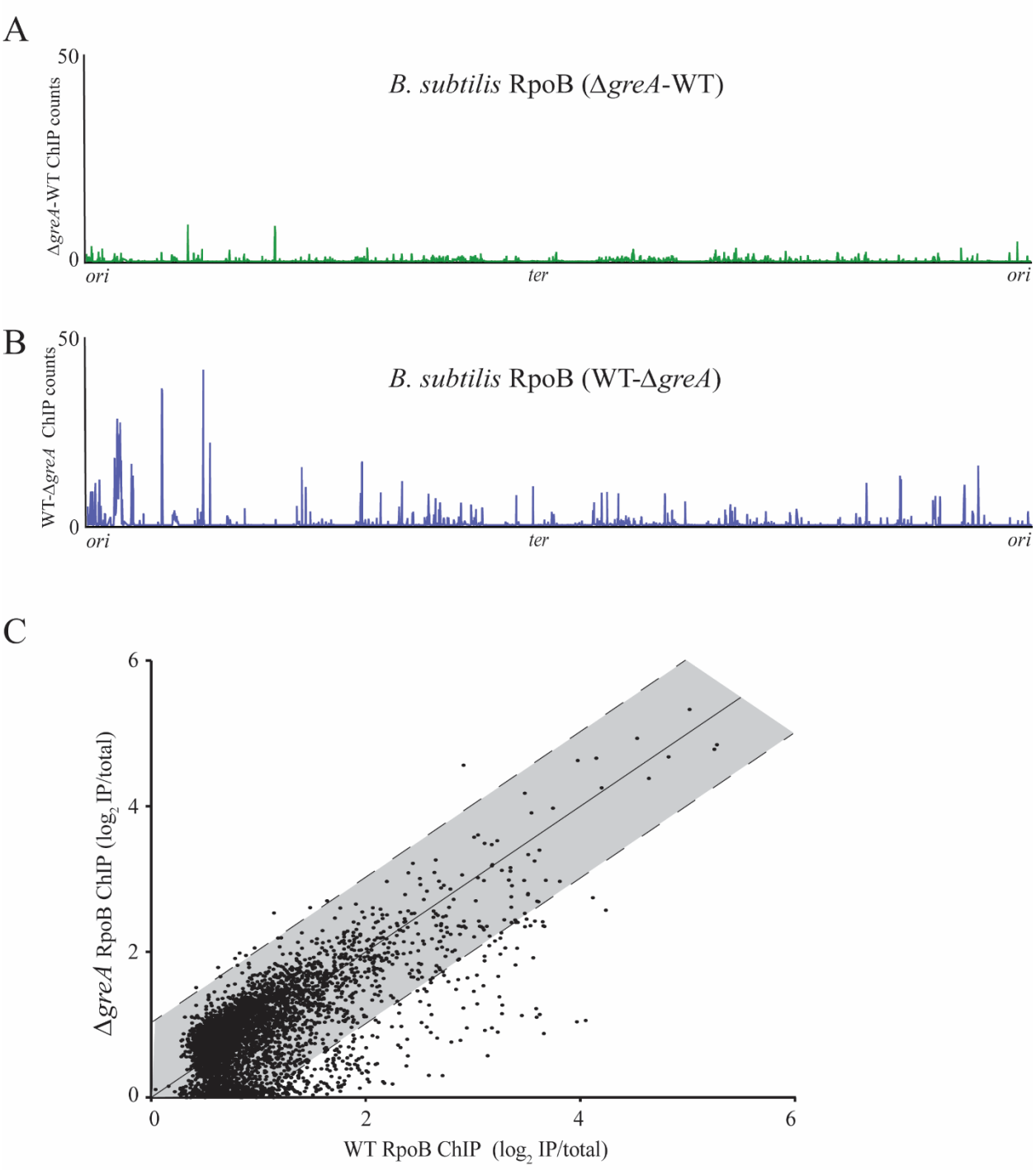

Figure 3.8 RNAP termination at structured regulatory RNAs is specific to Mfd
RpoB ChIP-seq plots showing regions of RpoB enrichment in *ΔgreA* relative to WT (A) and WT relative to *ΔgreA* (B). ChIP-seq read counts were normalized as described in Figure 3.5. (C) Scatter plot of WT and *ΔgreA* RpoB ChIP-seq. Quantification of ChIP signal was performed as described in Figure 2B.

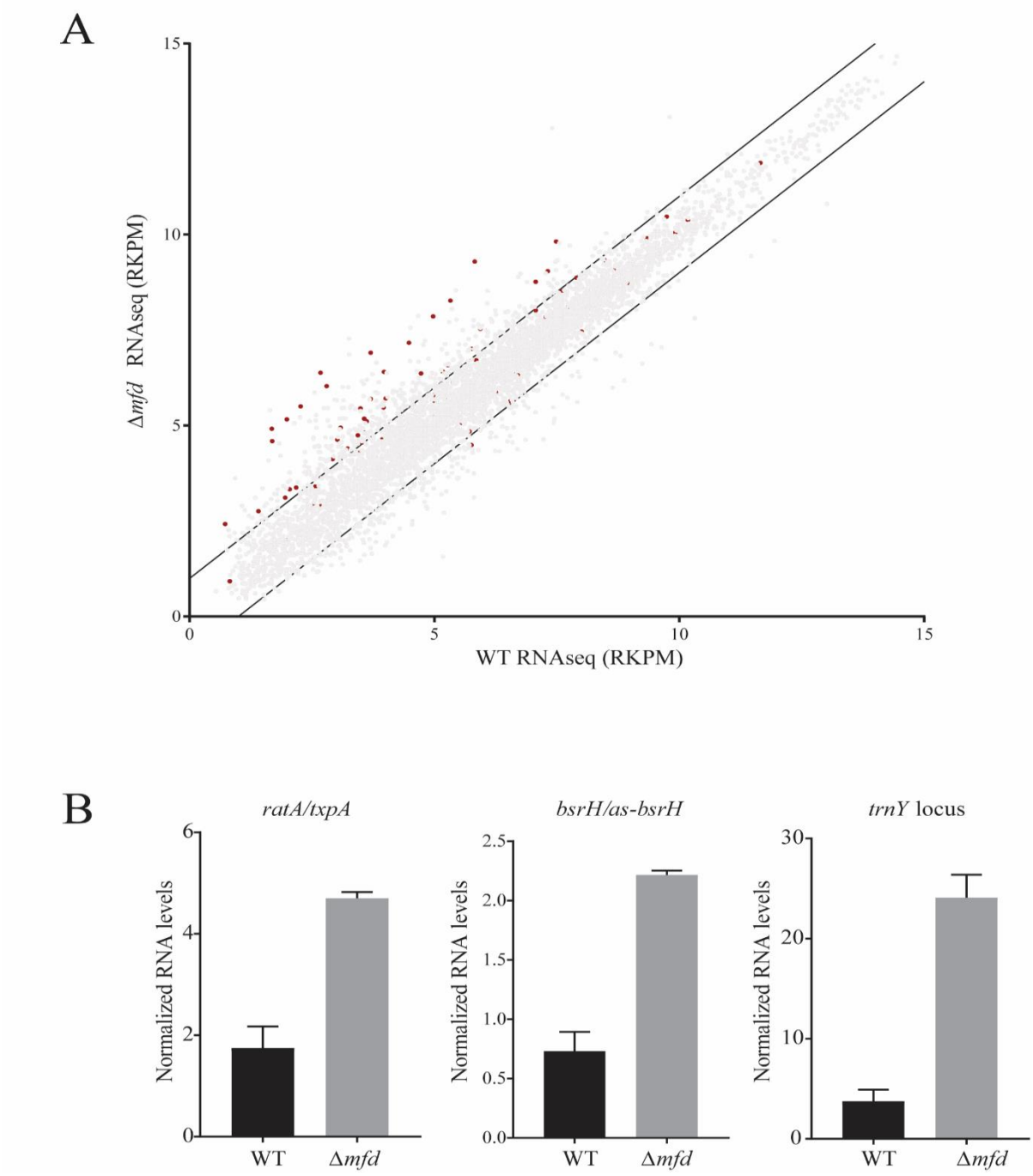

Figure 3.9 Mfd promotes repression of transcription at sites of structured sRNAs
(A) Scatter plot of RNA-seq. Data points represent the expression level of each gene in *B. subtilis* in WT and *Δmfd* strains. Scatter plot represents expression level calculated using read per kilobase per million reads (RPKM)[201], from at least two independent experiments. Data points in red indicate genes with increased RpoB occupancy in *Δmfd*. (B) qRT-PCR analysis of three regions with increased RNAP occupancy in *Δmfd txpA/ratA* (left), *bsrH/as-bsrH* (middle), and the *trnY* locus (right). RNA values normalized to ribosomal RNA.

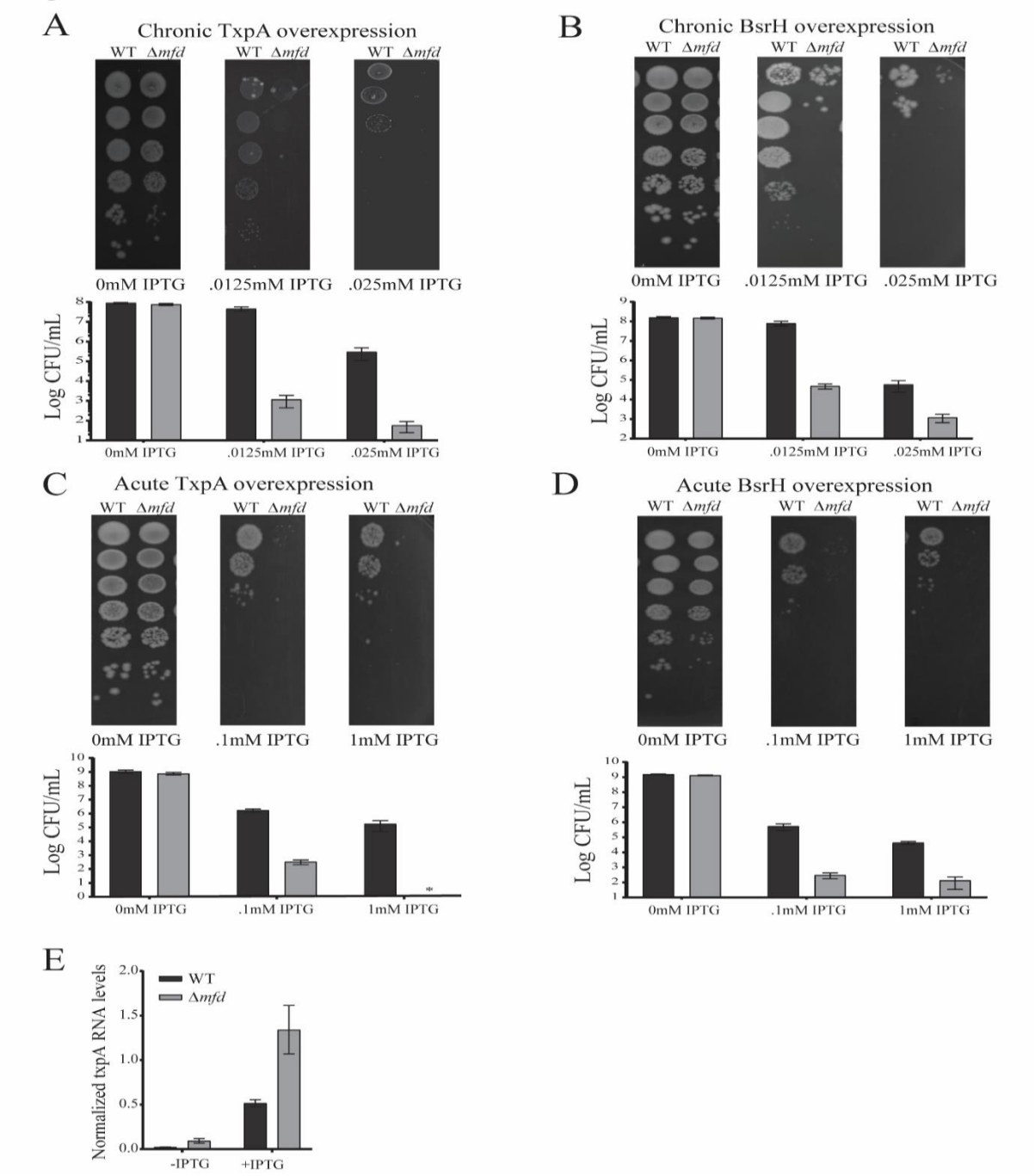

Figure 3.10 Transcriptional regulation by Mfd at toxin-antitoxin loci is essential for cell survival

Survival assays under chronic (A) and acute (B) overexpression of TxpA toxin in WT and *Δmfd* and survival assays under chronic (C) and acute (D) overexpression of BsrH toxin in WT and *Δmfd*. For all figures, representative images shown above, and quantification of data shown below. Error bars represent the S.E.M. from at least three independent experiments. (E) qRT-PCR analysis of txpA overexpression in WT and *Δmfd* strains. RNA values normalized to ribosomal RNA

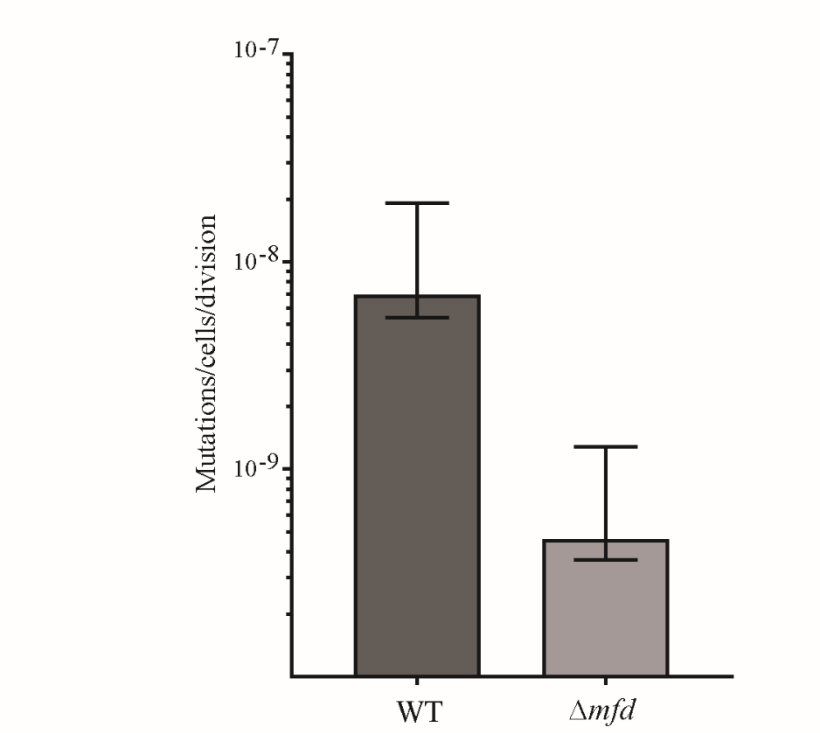

Figure 3.11 Mfd promotes mutagenesis at toxin-antitoxin loci
Mutation rate analysis of WT (black) and of *Δmfd* (grey) strains containing (A) ectopic TxpA overexpression ($P_{spank(hy)}$-*txpA*) or (B) ectopic BsrH overexpression ($P_{spank(hy)}$-*bsrH*). At least 24 replicates were performed for each strain, from at least three independent experiments. Error bars are 95% confidence intervals.

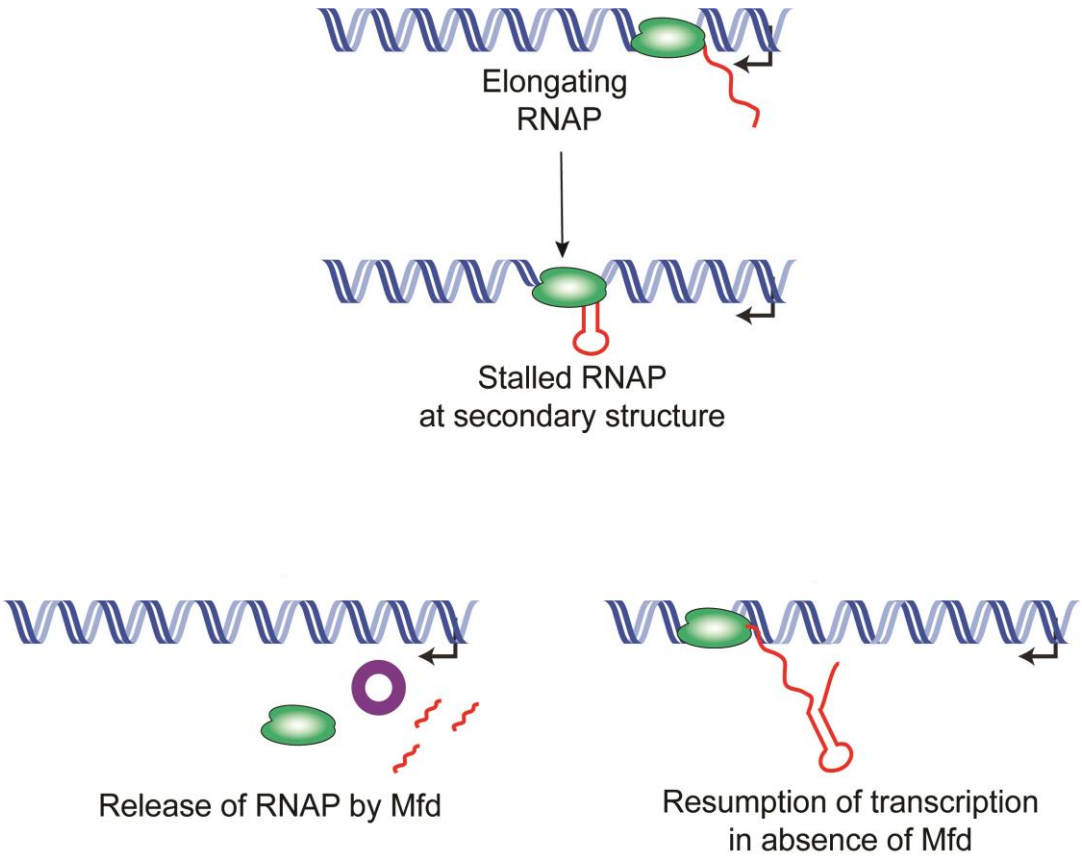

Figure 3.12 Model of Mfd activity at structured regulatory RNAs
During transcription, elongating RNAP (shown in green) transcribed a highly structured RNA sequence (shown in red). This can arrest RNAP (shown in maroon) on DNA. Bottom left- the arrested complex is recognized by Mfd (shown in purple), which releases RNAP from the template and in doing so represses transcription and promotes mutagenesis (shown as yellow stars on DNA). Bottom right- in the absence of Mfd, RNAP resumes transcription, leading to higher levels of RNA and decreased mutagenesis.

	Increased RpoB density in *Δmfd*	Decreased RpoB density in *Δmfd*
Genes (and TUs)	116 (71)	53 (31)
Genes (and TUs) bound by Mfd	52 (35)	0 (0)

Table 3.1 Summary list of genes and previously defined TUs with changes in RpoB density in *Δmfd* from ChIP-seq analysis
Changes in RpoB density and Mfd binding at TUs are defined by changes in one or more genes corresponding to its associated TUs

Increased RpoB density in *Δmfd*	logFC	logCPM	p-value	FDR
yjdF	4.240143	8.478652	3.02E-50	8.69E-47
S439	4.550725	7.570538	2.07E-43	3.98E-40
manA	4.416186	10.78631	3.89E-35	4.48E-32
yezE	2.840991	8.590838	3.77E-31	3.62E-28
yesE	2.807884	7.636147	2.6E-26	2.14E-23
yjdG	3.166576	7.065395	2.9E-25	2.09E-22
ybfA	2.37738	8.18114	1.22E-21	7.05E-19
ybeF	2.981832	7.020102	1.84E-20	9.63E-18
S440	3.991685	6.54651	4.16E-20	1.99E-17
manP	3.336177	11.42917	2.26E-19	1E-16
txpA	1.883679	8.770592	4.63E-17	1.78E-14
ilvE	1.874493	8.856673	2E-16	6.78E-14
bsrH	1.824349	8.894313	3.89E-16	1.16E-13
yxaD	2.128297	7.467853	4.02E-16	1.16E-13
ybgB	2.015781	7.986074	5.63E-16	1.54E-13
S1538	2.784014	5.92024	1.02E-15	2.56E-13
S978	1.73916	9.24128	1.16E-15	2.79E-13
S976	1.787689	8.910661	2.92E-15	6.72E-13
S977	1.695577	9.685895	9.35E-15	2.07E-12
S92	2.387837	6.449696	2.04E-14	4.36E-12
gudD	1.777336	8.315699	3.34E-14	6.62E-12
yxzK	2.028991	7.262953	5.92E-14	1.14E-11
S782	1.9941	7.269594	6.43E-14	1.19E-11
S345	1.94357	8.870044	3.37E-13	5.38E-11
yczL	2.47948	5.828152	6.38E-13	9.92E-11
S823	2.485837	5.386668	3.07E-12	4.11E-10
S346	1.843875	8.769623	4.63E-12	5.92E-10
ywnA	2.079491	6.847896	6.89E-12	8.63E-10
yabE	1.478489	11.43061	9.73E-12	1.19E-09
mccA	1.726044	7.846734	1.83E-11	2.15E-09
gudP	1.548927	8.094453	2.51E-11	2.84E-09
S979	1.508321	8.599922	2.8E-11	3.09E-09
maeA	1.512883	8.555576	3.32E-11	3.54E-09
S1553	1.651975	8.582116	5.13E-11	5.37E-09
S25	1.396479	11.53116	7.6E-11	7.57E-09
S654	2.9683	4.186248	7.63E-11	7.57E-09
S655	2.959525	4.260392	7.93E-11	7.73E-09
yhfP	1.446878	8.484789	1.92E-10	1.78E-08
mhqA	1.510673	7.648047	2.13E-10	1.91E-08
S653	2.934049	4.054978	2.13E-10	1.91E-08
bsrG	1.470846	7.822464	2.45E-10	2.14E-08

S438	2.206408	6.45237	2.69E-10	2.31E-08
yjdH	1.879546	6.580558	3.01E-10	2.54E-08
S364	2.025743	6.923438	4.36E-10	3.64E-08
ywnB	1.701565	7.061756	7.03E-10	5.7E-08
S810	1.5058	8.268258	8.15E-10	6.42E-08
S809	1.470097	8.299551	1.17E-09	8.88E-08
S1539	1.417572	8.026754	2.18E-09	1.57E-07
yhaX	1.526212	8.803472	2.33E-09	1.65E-07
rnmV	1.349754	8.745399	2.7E-09	1.9E-07
S186	1.92546	5.63386	6.9E-09	4.56E-07
aseR	2.022344	6.022738	7.76E-09	5.07E-07
S783	1.837128	5.99521	8.54E-09	5.52E-07
pghL	1.837595	5.809915	1.55E-08	9.84E-07
S441	2.404661	5.420373	1.6E-08	9.99E-07
S83	1.990035	5.300313	1.63E-08	9.99E-07
ilvD	1.272791	8.302726	2.75E-08	1.64E-06
S81	1.612237	6.850466	3.4E-08	1.99E-06
S1123	1.257309	8.608	4.16E-08	2.42E-06
cydC	1.184826	9.860062	5.06E-08	2.91E-06
abnA	1.576335	6.734966	5.79E-08	3.27E-06
yczJ	1.291698	7.896862	6.81E-08	3.8E-06
S492	1.777939	6.307327	9.51E-08	5.16E-06
yvdA	1.582491	6.67843	1.13E-07	6.09E-06
S136	1.256641	8.023684	1.26E-07	6.72E-06
S135	1.302045	7.600075	1.62E-07	8.55E-06
S310	1.82328	5.492673	1.81E-07	9.4E-06
yesK	1.620105	6.322324	2.31E-07	1.19E-05
walR	1.145347	9.037687	2.41E-07	1.23E-05
yesL	1.433417	6.880814	2.52E-07	1.27E-05
manR	2.145556	11.0353	3.16E-07	1.57E-05
yhaZ	1.222112	7.830292	3.3E-07	1.6E-05
S163	1.108061	8.963607	3.31E-07	1.6E-05
S811	1.270951	7.844985	5.01E-07	2.38E-05
yczI	1.36488	7.102516	5.46E-07	2.58E-05
S1554	1.298812	10.49104	6.91E-07	3.23E-05
S82	1.945362	4.92547	8.46E-07	3.89E-05
S214	1.370396	6.815781	9.72E-07	4.4E-05
S1434	2.061917	4.294567	1.18E-06	5.21E-05
ndoA	1.047178	9.153942	1.39E-06	6.05E-05
S1552	1.380252	7.194486	1.42E-06	6.14E-05
yugG	1.243201	7.308956	1.64E-06	7.03E-05
S1087	2.069925	5.07768	2.05E-06	8.62E-05
yndK	1.63855	5.055146	2.17E-06	9.04E-05

tdh	1.199971	7.753958	2.6E-06	0.000107
yesF	1.251569	7.36637	2.67E-06	0.000109
mntA	1.004916	9.808912	2.69E-06	0.000109
S460	1.66529	4.817007	3.43E-06	0.000135
atpI	1.211649	10.2215	3.48E-06	0.000135
yqxK	1.136915	7.625893	3.66E-06	0.000142
S1426	1.212438	9.614506	3.69E-06	0.000142
S1513	1.301223	7.694328	4.15E-06	0.000158
flvL	1.447298	5.942276	4.69E-06	0.000175
gntR	1.166278	7.802522	5.6E-06	0.000207
S787	1.652428	4.743454	6.51E-06	0.000239
S27	1.374183	6.725464	6.62E-06	0.000241
lcfB	1.021595	8.329309	6.87E-06	0.000246
glpF	1.342049	8.466848	6.9E-06	0.000246
dctP	1.133068	8.537037	6.96E-06	0.000246
ydfN	1.383896	6.536061	6.97E-06	0.000246
S321	1.245724	7.488175	9.04E-06	0.000311
gamR	1.132743	7.350659	9.17E-06	0.000314
ygaN	1.254975	6.717372	9.62E-06	0.000326
ypgR	1.023024	7.747455	9.9E-06	0.000333
S1203	1.419781	5.829353	1.14E-05	0.000381
wprA	1.016197	8.894388	1.22E-05	0.000403
spoVM	1.40264	6.409592	1.35E-05	0.000444
desE	1.20561	6.805016	1.5E-05	0.000488
sda	1.519732	5.185282	1.59E-05	0.000516
zinU	1.035096	7.504286	1.61E-05	0.000518
S1537	1.481547	5.434969	1.79E-05	0.000572
yqjY	1.383191	5.907056	1.87E-05	0.000594
thiN	1.118594	7.299829	1.93E-05	0.000605
S730	1.783409	4.06811	2.26E-05	0.000699
glpT	1.008189	7.460125	2.91E-05	0.000885
S895	1.431985	6.20782	3.01E-05	0.000906

Table 3.2 Genes with increased RpoB ChIP association in Δ*mfd*
List is sorted by increasing p-value. (logFC= log-fold change, logCPM= log counts per million, FDR= false discovery rate). The following criteria were used to define increased RpoB association= logFC>1, logCPM>4, p-value< $1x10^{-4}$, FDR< .001.

Decreased RpoB density in *Δmfd*	logFC	logCPM	p-value	FDR
srfAA	-2.26485	12.22756	3.90632E-19	1.61E-16
rsbV	-1.92267	9.001644	1.22689E-16	4.41E-14
rsbW	-1.78834	9.416076	3.17344E-16	1.01E-13
srfAB	-2.16679	12.03301	7.17914E-16	1.88E-13
Dps	-1.87902	8.060913	2.40446E-14	4.94E-12
csbA	-1.99084	7.22279	1.06434E-13	1.86E-11
comS	-2.35255	6.993624	1.75936E-13	2.98E-11
sigB	-1.66845	9.123723	2.11423E-13	3.48E-11
Hpf	-1.50878	10.14458	2.09592E-12	2.95E-10
rsbX	-1.66793	8.025089	2.14E-12	2.95E-10
srfAC	-1.79903	10.33624	2.15202E-12	2.95E-10
ytxG	-1.6182	7.720172	2.34443E-11	2.7E-09
ywjC	-2.11358	6.387004	2.98547E-11	3.24E-09
mtlF	-2.26126	8.058263	5.84657E-11	6.01E-09
yjbC	-1.44499	8.622561	1.44742E-10	1.37E-08
srfAD	-1.71142	7.540153	6.61549E-10	5.44E-08
mgsR	-1.88625	6.890448	8.02517E-10	6.41E-08
yjzE	-1.77142	6.824531	8.35835E-10	6.5E-08
ctc	-1.37246	10.88936	1.24184E-09	9.16E-08
clsC	-1.45099	8.674811	1.71788E-09	1.25E-07
S408	-1.67932	6.807454	3.92146E-09	2.72E-07
csbB	-1.37285	8.049917	4.0486E-09	2.77E-07
ypiB	-1.5183	7.316947	5.043E-09	3.41E-07
ahpC	-1.25127	9.245092	5.61543E-09	3.76E-07
ybeC	-1.30304	9.616783	1.55568E-08	9.84E-07
sunT	-1.55813	7.379107	2.11823E-08	1.28E-06
ybyB	-1.60283	7.149778	5.41157E-08	3.08E-06
S1343	-1.64551	6.436233	7.81885E-08	4.33E-06
ytxH	-1.43845	7.151014	9.41537E-08	5.16E-06
rsbU	-1.17501	8.566677	1.64451E-07	8.6E-06
ptsG	-1.0742	9.870055	3.21753E-07	1.58E-05
pdaC	-1.49401	8.022321	4.868E-07	2.33E-05
S928	-1.46714	6.647883	7.49756E-07	3.48E-05
S1446	-1.27931	7.218121	9.03486E-07	4.13E-05
gtaB	-1.05855	8.909825	1.06347E-06	4.78E-05
ahpF	-1.07837	10.72555	1.13678E-06	5.07E-05
S476	-1.44632	6.386425	1.26638E-06	5.56E-05
serA	-1.0807	8.50521	1.82054E-06	7.76E-05
S1342	-1.65166	5.177954	1.85905E-06	7.87E-05
ywjB	-1.14097	7.505803	2.23406E-06	9.25E-05
S927	-1.26984	7.173581	3.25117E-06	0.00013
S1366	-2.08333	4.517936	3.42136E-06	0.000135

yjbB	-1.13823	7.621385	3.48353E-06	0.000135
icd	-1.06232	8.432157	4.47548E-06	0.000169
S294	-1.84262	4.333745	4.62556E-06	0.000174
S1171	-1.61913	4.899514	5.05606E-06	0.000188
S427	-1.529	5.39853	6.91537E-06	0.000246
yjcF	-1.24114	6.860261	8.01028E-06	0.000279
mtlD	-1.39012	8.150005	1.37677E-05	0.00045
S1301	-1.03971	8.236248	1.87937E-05	0.000594
S426	-1.37114	5.683764	2.1616E-05	0.000676
rsbRD	-1.25851	6.778225	2.6172E-05	0.000805
ykoL	-1.36654	5.951867	2.65882E-05	0.000814

Table 3.3 Genes with decreased RpoB ChIP association in Δ*mfd*
Genes are sorted by p-value. Criteria used to define decreased RpoB association is the same as described in Table 3.2.

TUs with Mfd binding and increased RpoB in *Δmfd*	regulatory RNA category
manP-manA-S439-yjdF	riboswitch, intergenic
S442-yjdH-S441-yjdG	5’ UTR, intergenic
S81-ybeF-ybfA-ybfB	5’ UTR
txpA (as ratA)	asRNA, ncRNA
bsrH (as bsrH)	asRNA, ncRNA
S782-S783-yopT	5’ UTR
S345-S346-yhaX	independent transcript
S823-ilvD	5’ UTR
yabE (S25 asRNA)	asRNA, ncRNA
yrrT-mtnN-S1033-mccA-mccB-yrhC	Intergenic
S1434-maeE	5’ UTR
S1552-S1553-walR-walK-walH-walI-walJ-htrC	5’ UTR
S655-S654	independent transcript
yhfO-yhfQ-S364-yhfP	Intergenic
S460-mhqA	5’ UTR
bsrG (SR4 asRNA)	asRNA, ncRNA
S438-yjdB-S437	5’ UTR, 3’ UTR
S27-rnmV-ksgA	5’ UTR
S1123-nifZ-thiI-sspA	5’ UTR
S1487-S1486-cydA-cydB-cydC-cydD	5’ UTR
S492-clpE	5’ UTR
manR	None
ndoAI-ndoA asRNA(S163-S164-S165)	asRNA
S811	independent transcript
trnY locus	tRNA, intergenic
alaR-alaT-S1201	3’ UTR
S1175-S1174-mntA-mntB-mntC-mntD	5’ UTR
S1427-S1426-atpI-atpB-atpE-atpF-atpH-atpA-atpG-atpD-atpC	5’ UTR
S895-yqxK	5’ UTR
S1513-bglP-bglH-yxiE	5’ UTR, riboswitch
S321-glpF-glpK	5’ UTR, riboswitch
mhqN-mhqO-mhqP	None
S1203-yugF	5’ UTR
spoVM	None
S966-sdA (asRNA s965)	5’ UTR, asRNA
polY2-yqjX-S898-S897-yqjY-yqjZ-yqkA-yqkB-yqkC	Intergenic
S82-S83-glpT-glpQ	5’ UTR, riboswitch

Table 3.4 TUs with Mfd binding and increased RpoB occupancy in *Δmfd*
Associated sRNA categories from previously defined work[202].

TUs with decreased RpoB in *Δmfd*	regulatory RNA category
srfAA-srfAB-comS-srfAC-srfAD	None
rsbR-rsbS-rsbT-rsbU-rsbV-rsbW-sigB-rsbX	None
S1343-csbA-S1342	5’ UTR, 3’ UTR
Hpf	None
ytxG-ytxH-ytxJ	None
ywjC-S1446	3’ UTR
mtlA-mtlF-mtlD	None
S408-yjbC-S409-spx	5’ UTR, intergenic
S928-mgsR	5’ UTR
S426-yjcD (asS427-yjzE)	asRNA
Ctc	None
ywiE-ywjA-ywjB	None
S294-csbB	5’ UTR
ypiA-ypiB	None
ahpF-ahpC	None
ybeC	None
sunA-sunT-bdbA-yolJ-bdbB	None
Ybyb	None
ptsG-ptsH-ptsI	None
pdaC	None
gtaB-S1363	3’ UTR
S476-ykoM	5’ UTR
serA	None
rsbRD (as927)	asRNA
S1366	independent transcript
yjbB	None
Icd	None
S1171-ytkA-S1172-dps	5’ UTR, intergenic
yjcH-yjcG-yjcF	None
S1301	independent transcript
ykzB-ykoL	None

Table 3.5 TUs with decreased RpoB association in *Δmfd*
Associated sRNA categories from previously defined work[202]

Strain	Mutation
WT1	Y4H (**T**AC -> **C**AC)
WT2	V9G (G**T**C -> G**G**C)
WT3	-65 (A -> G)
WT4	-65 (A -> G)
WT5	-69 (T->G)
WT6	111Δ1
WT7	-65 (A -> G)
WT8	-65 (A -> G)
Δmfd1	Y4H (**T**AC -> **C**AC)
Δmfd2	Y4H (**T**AC -> **C**AC)
Δmfd3	25Δ1
Δmfd4	25Δ1
Δmfd5	-65 (A -> G)
Δmfd6	-65 (A -> G)
Δmfd7	-65 (A -> G)
Δmfd8	-69 (T->G)

Table 3.6 Mutations conferring resistance to *txpA* overexpression
Mutations identified in 8 WT and 8 *Δmfd* strains from Luria-Delbruck fluctuation assays. Point Mutations are indicated by the amino acid substitution and nucleotide substitution for protein coding changes. Deletions are indicated by location within the *txpA* locus and the size of the deletion and upstream mutations are indicated solely by the nucleotide substitution

Chapter 4

Concluding remarks

5.1 Summary of findings

Through my work we have discovered that Mfd is a critical factor for the rapid rise of AMR. Mfd-mediated evolution is a highly conserved phenomenon, that occurs in divergent gram-positive and gram-negative bacterial pathogens, as well as in *Mtb*. This phenomenon is also relevant to diverse classes of antibiotics, with completely different mechanisms of action. This work suggests that mutagenesis promoted by Mfd is an inherent, fundamental process, and thus Mfd can be thought of as a highly conserved evolvability factor in all cells. We therefore propose that targeting the activity of Mfd with an inhibitor may reduce the ability of bacterial pathogens to evolve clinical drug resistance. More broadly, the identification and ultimate inhibition of pathways the promote mutagenesis and evolution in bacteria may lead to a new class of therapeutics that could help curtail the rise of AMR.

I next turned my attention to more deeply describing endogenous regions that may be prone to Mfd-mediated mutagenesis, and in doing so, unraveled novel mechanistic insights into Mfd's *in vivo* functions. ChIP-seq studies of Mfd and RpoB with and without Mfd revealed that Mfd functions as an RNAP co-factor, and its co-factor activity is particularly critical at highly structured regulatory RNAs. At these sites, I discovered that Mfd promotes release of RNAP and in doing promotes transcriptional repression. Regulation of transcription by Mfd at highly structured regulatory RNAs is critical for cell viability, as in its absence, cells are highly sensitized to expression of toxin-antitoxin systems containing structured RNAs. Lastly, I find

that Mfd promotes mutagenesis at toxin-antitoxin loci which express structured regulatory RNAs. These findings link Mfd activity and Mfd-mediated mutagenesis to the evolution of structured regulatory RNAs.

5.2 Future directions

Much work is needed to more deeply understand the mechanism by which Mfd promotes mutagenesis, what other regions of the genome are particularly prone to mutagenesis, and how Mfd's activity at site expressing high RNA secondary structure may fundamentally promote diversity at both the DNA and RNA level.

The underlying mechanism of Mfd induced mutations

Precisely how Mfd promotes mutagenesis remains unclear. Various studies have proposed a wide range of models[85,87,150,203], yet a definitive experiment has yet to be performed. This is largely because traditional assays that measure mutagenesis are quite restrained in their ability to resolve mechanism. Classic reversion assays that measuring rates of antibiotic resistance or measure reversion rates in the context of engineered reporters, only capture one (or a few) critical mutagenic events. Very little information, if any, can be gathered about the diverse range of mutations that truly arise in a bacterial population during growth. The solution to this problem is to measure mutagenesis in that absence of selection, but given the inherently low rate of mutation, such approaches are technically challenging.

Mutation accumulation (MA) assays coupled with whole-genome sequencing in one approach to identify natural mutations that arise in the absence of selection[204]. Such experiments require

serial bottlenecking and leverage the phenomenon of genetic drift to fix random mutations in populations, in the absence of selection. Such assays can be informative but require thousands of generations of growth to accumulate a very small number of mutations, and therefore suffer from challenges relating to low resolution.

Recent advances in deep sequencing technology allow us measure mutation rates, in the absence of selection, at specific loci with significantly higher resolution[52,205]. Harnessing these approaches will allow us to capture rare mutational events that occur subpopulations of cells, gathering a far more complex and nuanced picture of mutagenesis. These technologies will help guide our mechanistic understanding of mutagenesis in future studies. Importantly, applying these approaches to our work will allows us to gain mechanistic insight into Mfd-mediated mutagenesis.

RNA secondary structure and mutagenesis

A second unresolved area pertaining to my work is the need to gain a more detailed understanding of how RNA secondary structure promotes Mfd-mediated mutagenesis. Our work suggests that RNA secondary structure may be a novel category of transcription-associated mutagenesis (TAM), and that Mfd plays a role in this phenomenon, but experiments directly testing this hypothesis will be deeply informative. Utilizing controlled reporter constructs, where we deliberately and rigorously alter the presence and the extent of RNA secondary structures, and then measure mutagenesis, will provide a better understanding of how RNA structure influences mutagenesis and evolution. We can also harness the deep-sequencing technologies previously discussed to identify the mutations that arise at endogenous DNA regions that express

structured RNAs, such as tRNA loci or 5’ UTR regions containing riboswitches. Such experiments would allow us to better understand the nature of mutations that are promoted at these sites, and how they may differ from other regions in the genome. These additional experiments could excitingly lead to an undiscovered mechanism of mutagenesis.

Chapter 5

Materials and Methods

5.1 Strains and plasmids built and used in these studies

Strain number	Genotype	Source/details of strain construction
HM1	WT *B. subtilis* JH642	Merrikh Lab stock
HM713	*B. subtilis* 168 *ΔuvrA*::mls	From *B. subtilis* Genetic Stock Center
HM1720	*B. subtilis* 168 *Δmfd*::mls	From *B. subtilis* Genetic Stock Center
HM1996	WT *S.typhimurium* ST19	Miller Lab stock- clinical isolate
HM2212	WT *P. aeruginosa* CF127	Parsek Lab stock- cystic fibrosis isolate
HM2260	*P. aeruginosa* CF127 Δ*mfd*	Transformation of pHM661 into *P. aureginosa* CF127
HM2295	*E. coli* F' (Kan) placOL2–62-lacZ	Gift from Dr. Ann Hochschild
HM2521	*B. subtilis* JH642 Δ*mfd*::mls	Transformation of HM1720 genomic DNA into HM1
HM2602	*E. coli* F' (Kan) placOL2–62-lacZ pSIM27(tet)	Transformation of pSIM27 into HM2295 (Addgene 53735)
HM2633	B. subtilis JH642 Δ*uvrA*::mls	Transformation of *ΔuvrA* genomic DNA into HM1
HM2747	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg)	Recombineering into HM2602 using pSIM27. Electroporation of hygromycin-marked amplicon using primers HM3391 and HM3392 with pHM443 as template.
HM2838	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pACλCI(Cat) pHM458(Amp)	Transformation of pACλCI(Cat) and pHM458(Amp) into HM2747
HM2875	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pHM474(Cm) pBRα(Amp)	Transformation of pHM474 and pBRα(Amp) into HM2747
HM2880	WT *S.typhimurium* ST19 + pUC19	Transformation of pUC19 into HM1996
HM2881	*S. typhimurium* ST19 Δ*mfd*::cat + pUC19	Transformation of pUC19 into HM3429
HM2882	*S. typhimurium* ST19 Δ*mfd*::cat + pHM481	Transformation of pHM481 into HM3429
HM2886	*S. typhimurium* ST19 Δ*mfd*::cat + pHM484	Transformation of pHM482 into HM3429
HM2913	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pHM474(Cm) pHM480(Amp)	Transformation of pHM474 and pHM480 into HM2747
HM2920	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pHM453(Cm) pHM457(Amp)	Transformation of pHM453 and pHM457 into HM2747
HM2921	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pHM453(Cm) pHM494(Amp)	Transformation of pHM453 and pHM494 into HM2747
HM2925	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pHM453(Cm) pBRα(Amp)	Transformation of pHM453 and pBRα into HM2747
HM2926	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pACλCI(Cat) pHM457(Amp)	Transformation of pACλCI and pHM457 into HM2747
HM2949	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pACλCI(Cat) pHM480(Amp)	Transformation of pACλCI and pHM480 into HM2747

HM2962	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pHM499(Cm) pHM480(Amp)	Transformation of pHM480 and pHM499 into HM2747
HM3134	*S. typhimurium* ST19 Δ*mfd*::cat + pHM662	Transformation of pHM662 into HM3429
HM3193	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pSIM27(tet) pHM550(Cm) pBRα(Amp)	Transformation of pHM550(Cat) and pBRα(Amp) into HM2747
HM3225	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pSIM27(tet) pHM550(Cm) pHM458(Amp)	Transformation of pHM550(Cat) and pHM458(Amp) into HM2747
HM3245	*S. typhimurium* ST19 *ΔuvrA::*kan	Recombineering into HM1996 using pKD46. Electroporation of kanamycin marked amplicon using primers HM4064 and HM4065 and pKD13 as template
HM3406	*S. typhimurium* ST19 *Mtbmfd*-kan	Recombineering into HM1996 using pHM629. Electroporation of *Mtb-mfd* coding sequence with kanamycin marker downstream, using primers HM3632 and HM3635
HM3429	*S. typhimurium* ST19 *Δmfd::*cat	Recombineering into HM1996 using pKD46. Electroporation of chloramphenicol marked amplicon using primers HM2374 and HM2375 and pKD3 as template
HM3585	WT S.typhimurium ST19 + pMMB67EH	Transformation of pMMB67EH into HM1996
HM3586	*S. typhimurium* ST19 Δ*mfd*::cat + pMMB67EH	Transformation of pMMB67EH into HM3429
HM3590	*S. typhimurium* ST19 Δ*mfd*::cat + pHM649	Transformation of pHM649 into HM3429
HM3666	*S. typhimurium* ST19 Δ*mfd*::cat + pHM650	Transformation of pHM650 into HM3429
HM3667	*S. typhimurium* ST19 Δ*mfd*::cat + pHM651	Transformation of pHM651 into HM3429
MR01	WT *Mtb* H37Rv	Sherman lab stock
MR02	*Mtb* H37Rv *Δmfd*::hyg	Recombineering into MR01 using pNIT. Electroporation of hygromycin marked amplicon using primers HM2877 and HM3908 and pHM566 as template
pHM443	pKM342-Nanoluc	Built from pKM342 backbone (BglII digested) and Nanoluc amplicon (primers HM3202 and HM3203)
pHM453	pACλCI Plac-CI-ST19rpoB(19-142)	Built with Gibson cloning from pACλCI-β-flap backbone (BamHI/NotI digested) and RpoB(AA19-142) amplicon with stop codon added (primers HM3516 and HM3517)
pHM457	pBRα Plac-a-ST19mfd (RID)	Built with Gibson cloning from pBRα-β-flap backbone (BamHI/NotI digested) and Mfd RID amplicon with stop codon added (primers HM3512 and HM3513)
pHM458	pBRα Plac-a-ST19rpoB(19-142)	Built with Gibson cloning from pBRα-β-flap backbone (BamHI/NotI digested) and RpoB(AA19-142) amplicon with stop codon added (primers HM33518 and HM3519)

pHM474	pACλCI Plac-CI-ST19mfd(1-450)	Built with Gibson cloning from pACλCI-β-flap backbone (BamHI/NotI digested) and Mfd(1-450) amplicon with stop codon added (primers HM3506 and HM3546)
pHM480	pBRα Plac-a-ST19UvrA(88-505)	Built with Gibson cloning from pBRα-β-flap backbone (BamHI/NotI digested) and UvrA(88-505) amplicon with stop codon added (primers HM3690 and HM3691)
pHM481	pUC19 ST19 mfd	Built using restriction enzyme cloning. pUC19 backbone was amplified using primers HM3719 and HM3720 and then digested with Xho and Nhe. *S. typhimurium* mfd (including promoter region) was amplified using primers HM3717 and HM3718 and digested with Xho and Nhe.
pHM484	pUC19 ST19 mfd L499R	Built using site-directed mutagenesis of pHM481 using primers HM3725 and HM3726
pHM494	pBRα Plac-a-ST19mfd (RID) L499R	Built using site-directed mutagenesis using primers HM3725 and HM3726 and pHM457 as template.
pHM499	pACλCI Plac-CI-ST19mfdR165A	Built using site-directed mutagenesis using primers HM3729 and HM3730 and pHM474 as template.
pHM550	pACλCI Plac-CI-*Mtb*mfd(334-651)	Built with Gibson cloning from pACλCI-β-flap backbone (BamHI/NotI digested) and *Mtb* Mfd(334-651) amplicon with stop codon added (primers HM4007 and HM4011)
pHM566	pKM342-mfd	Built with 2-step Gibson Cloning. Linear amplicons of pKM342 backbone (primers HM3045 and HM3046) and a 500bp homology region upstream Mfd locus (primers HM3047 and HM3048) were ligated using Gibson cloning. This intermediate plasmid and a 500bp homology region in the 3' end of the Mfd locus were linearized with primers HM3049 and HM3050 and HM3702 and HM3703, respectively
pHM629	pUC19- *Mtbmfd*-kan	SOE PCR was used to generate a linear amplicon of *Mtb*-mfd (primers HM3632 and HM3633) with Kanamycin resistance casette downstream (primers 3634 and 3635). Gibson cloning of SOE product and pUC19 amplicon (primers HM3646 and HM3647) was used to construct plasmid

pHM649	pMMB67EH_ST19mfd	Built with Gibson Cloning from linear amplicons of pMMB67EH backbone (primers HM4759 and HM4760) and Mfd (primers HM4761 and HM4762 and pHM481 as template. Contains native Mfd promoter)
pHM650	pMMB67EH_ST19 mfd R165A	Built using site-directed mutagenesis using primers HM3729 and HM3730 and pHM649 as template.
pHM651	pMMB67EH_ST19 mfd L499R	Built with Gibson Cloning from linear amplicons of pMMB67EH backbone (primers HM4759 and HM4760) and Mfd (primers HM4761 and HM4762 and pHM484 as template. Contains native Mfd promoter)
pHM661	pEX18-mfd	Built with SOE PCR of upstream and downstream homology regions of *P. aureginosa* Mfd (primers HM2577, HM2578 and HM2579, HM2580), and and restriction digest of amplicon and pEX18 plasmid with HindIII and BamHI
pHM662	pUC19 *Mtbmfd*	Built using Gibson cloning. pUC19 backbone containing ST19 Mfd promoter region was amplified using primers HM3646 and HM3647 and *Mtb* mfd coding sequence was amplified using SOE PCR (primers HM3692-HM3698 for full length Mtb mfd sequence), and primers HM3648 and HM3649 for Gibson homology
pBRα	Used as a negative control	Addgene 53731
pBRα-β-flap	Used to clone and express RNA polymerase α-subunit fusions in *E. coli*	Addgene 53734
pACλCI	Used as a negative control	Addgene 53730
pACλCI-β-flap	Used to clone and express λCI fusions in *E. coli*	Addgene 53733
pEX18	used for deletion construct in *P. aeruginosa*	Parsek Lab stock
pKD3	Cat cassette flanked by FRT sites for recombineering; template plasmid	Addgene 45604
pKD13	Kan cassette flanked by FRT sites for recombineering; template plasmid	Miller Lab stock
pKD46	mini λ Red recominbase expression plasmid for recomineering in *S. typhimurium*	Miller Lab stock
pMMBEH67	ampR- used as an empty vector control	Addgene 37622

pNIT	used for recombineering in *Mtb*	Sherman lab stock
pNL1.1	NanoLuc expression vector	Promega
pSIM27	used for recombineering in *E. coli*	A gift from the Court lab- https://redrecombineering.ncifcrf.gov/strains--plasmids.html
pUC19	ampR- used as an empty vector control	Addgene 50005

5.2 Primers used in these studies

Primer Number	**Oligo Sequence**	**Primer details**
HM2374	5'-AAGAGTGCGGCGTAAAACAAAAAGAGATACTGACAACCGTTGTGTAGGCTGGAGCTGCTTC -3'	amplification of pKD3 template to make mfd deletion in *S. typhimurium*
HM2375	5'-TAGGCCGGACGCAGAGTAATGACGCCCGGCCTGACGCTTACATATGAATATCCTCCTTAG -3'	amplification of pKD3 template to make mfd deletion in *S. typhimurium*
HM2376	5'- CAACGATCGCGGATAATAAAGAAC -3'	PCR confirmation of *S. typhimurium* mfd deletion
HM2377	5'- CTGGAACTGCAGGGAAGTG -3'	PCR confirmation of *S. typhimurium* mfd deletion
HM2577	5'- ATCCGGAAGCTTCTTCGTCGATGGTGATGGT -3'	amplification of locus upstream of *P. aureginosa* Mfd for cloning into pEX18
HM2578	5'-TAGGCTAATCCTTACACGGGCAAGCTGGGCTCCGGTCG -3'	amplification of locus upstream of *P. aureginosa* Mfd for cloning into pEX18
HM2579	5'-CCAGCTTGCCCGTGTAAGGATTAGCCTATGCGCTTGTT -3'	amplification of locus downstream of *P. aureginosa* Mfd for cloning into pEX18

HM2580	5'- ATCCGGGGATCCCGAAATCGAACAGGCGTTG -3'	amplification of locus downstream of *P. aureginosa* Mfd for cloning into pEX18
HM2581	5'- GTTCTTGATGCCGTACTGG -3'	PCR confirmation of *P. aureginosa* mfd deletion
HM2582	5'- GATATCGTCGGCATGACCG -3'	PCR confirmation of *P. aureginosa* mfd deletion
HM2831	5'- CGTCAGCCTGAAGTGAAAGAAG -3'	PCR and sequencing primer for confirming inserts in pBRα
HM2832	5'- ATCGGTGATGTCGGCGATATAG -3'	PCR and sequencing primer for confirming inserts in pBRα
HM2877	5'- CCTGTACGCGGTGGTGGTC -3'	amplification of hygromycin marked cassette (pHM566) to make Mfd deletion in *Mtb*
HM3045	5'- GTGGCGGCCGCTCTAGAACT -3'	amplification of pKM342 for construction of pHM566 (pKM342-mfd plasmid)
HM3046	5'- CGCGGTGGAGCTCCAATTCG -3'	amplification of pKM342 for construction of pHM566 (pKM342-mfd plasmid)
HM3047	5'- CGAATTGGAGCTCCACCGCGCCTGTACGCGGTGGTG GTC -3'	amplification of 5' UTR of *Mtb* mfd for construction of pHM566 (pKM342-mfd plasmid)
HM3048	5'- AGTTCTAGAGCGGCCGCCACGATGCGGCCATTCTAG GGC -3'	amplification of 5' UTR of *Mtb* mfd for construction of pHM566 (pKM342-mfd plasmid)

HM3049	5'- TTCTAGACTCGAGGTACCG -3'	linearization of pKM342 plasmid containing 5' UTR region of *Mtb* mfd for construction of pHM566
HM3050	5'- GTACTTAATTCGAGCTCGGT -3'	linearization of pKM342 plasmid containing 5' UTR region of *Mtb* mfd for construction of pHM566
HM3202	5'- ACGACGGCCAGTGAATTACTTAAGAATGGTCTTCAC ACTCGAAGATTTCGTTG -3'	amplification of nanoluc plasmid (pNL1.1) for construction of pHM443
HM3203	5'- ATAGTGAGTCGGCGCGCCACTAGTATTACGCCAGAA TGCGTTCGC -3'	amplification of nanoluc plasmid (pNL1.1) for construction of pHM443
HM3389	5'- TCGCTAGTCAGTGGCCTGAAG -3'	PCR and sequencing primer for confirming inserts in pACyCI
HM3390	5'- GGTAGCCAGCAGCATCCTG -3'	PCR and sequencing primer for confirming inserts in pACyCI
HM3391	5'- ACCGAGCGGATAACAATTTCACACAGGAAACAGCTA TGGTCTTCACACTCGAAGATTTCG -3'	amplification of hyg marked nanoluc plasmid (pHM443) for bacterial 2-hybrid system
HM3392	5'- GCGAAATACGGGCAGACATGGCCTGCCCGGTTATTA TCAGGCGCCGGGGG -3'	amplification of hyg marked nanoluc plasmid (pHM443) for bacterial 2-hybrid system
HM3443	5'- GTCTCCTCGAACACCTCGAA -3'	PCR confirmation of *Mtb* mfd deletion - binds to hyg cassete
HM3444	5'- AGAGCACCAACCCCGTACT -3'	PCR confirmation of *Mtb* mfd deletion - binds to hyg cassete

HM3446	5'- GTGGTCGATCGGGAGATGT -3'	PCR confirmation of *Mtb* mfd deletion - binds upstream of Mfd
HM3447	5'- GACGTCGTCGATGGTGAAC -3'	PCR confirmation of *Mtb* mfd deletion - binds upstream of Mfd
HM3506	5'-AGTGGCCTGAAGAGACGTTTGGCGCAGCGGCCGCAATGCCTGAAAAACAACGTTACAC -3'	amplification of *S. typhimurium* Mfd (AA 1-450) for bacterial 2-hybrid
HM3512	5'-AAGTGAAAGAAGAGAAACCAGAGGCAGCGGCCGCACGCACGATCAACCCGGATAC -3'	amplification of *S. typhimurium* Mfd (RID domain) for bacterial 2-hybrid
HM3513	5'-CGGCCACGATGCGTCCGGCGTAGAGTTAGAATAACTGATACTGCTCACGATCATGC -3'	amplification of *S. typhimurium* Mfd (RID domain) for bacterial 2-hybrid
HM3516	5'-AGTGGCCTGAAGAGACGTTTGGCGCAGCGGCCGCACCACAAGTTCTGGATGTACCT -3'	amplification of *S. typhimurium* rpoB (AA 19-142) for bacterial 2-hybrid
HM3517	5'-CTGCGATGCAGATCTGTAAGGTAAGTTACTCAGTACCGTTGATAACAAAGG -3'	amplification of *S. typhimurium* rpoB (AA 19-142) for bacterial 2-hybrid
HM3518	5'-AAGTGAAAGAAGAGAAACCAGAGGCAGCGGCCGCACCACAAGTTCTGGATGTACCT -3'	amplification of *S. typhimurium* rpoB (AA 19-142) for bacterial 2-hybrid
HM3519	5'-CGGCCACGATGCGTCCGGCGTAGAGTTACTCAGTACCGTTGATAACAAAGG -3'	amplification of *S. typhimurium* rpoB (AA 19-142) for bacterial 2-hybrid

HM3546	5'-CTGCGATGCAGATCTGTAAGGTAAGTTACAAAAGATCGCTTTCACAAATC -3'	amplification of *S. typhimurium* Mfd (AA 1-450) for bacterial 2-hybrid
HM3632	5'-AGAGTGCGGCGTAAAACAAAAAGAGATACTGACAACCGTTATGACCGCACCGGGGCC -3'	amplification of *Mtb* Mfd coding sequence with upstream homology to ST19 Mfd
HM3633	5'-GCCCCAGCTGGCAATTCCGGTCACGGTTGTCGCTCCTTAAT -3'	amplification of *Mtb* Mfd coding sequence with homology to Kan resistance casette from pKD13
HM3634	5'-ATTAAGGAGCGACAACCGTGACCGGAATTGCCAGCTGGGGC -3'	amplification of Kan resistance casette from pKD13 with homology to *Mtb* Mfd coding sequence
HM3635	5'-GATTAGGCCGGACGCAGAGTAATGACGCCCGGCCTGACGCTTAGAAGAACTCGTCAA -3'	amplification of Kan resistance casette from pKD13 with downstream homology to ST19 Mfd
HM3646	5'- AACGGTTGTCAGTATCTCTTTTTG -3'	amplification of pUC19 containing ST19 Mfd promoter region
HM3647	5'- CTCGAGTGGGGTGCCTAATG -3'	amplification of pUC19 containing ST19 Mfd promoter region
HM3648	5'-CAAAAAGAGATACTGACAACCGTTATGACCGCACCGGGGCCTG -3'	amplification of *Mtb* mfd for gibson cloning into pUC19
HM3649	5'-CATTAGGCACCCCACTCGAGTCACGGTTGTCGCTCCTTAATC -3'	amplification of *Mtb* mfd for gibson cloning into pUC19

HM3690	5'-AAGTGAAAGAAGAGAAACCAGAGGCAGCGGCCGCAAAATCGACATCGCACAACC -3'	amplification of *S. typhimurium* UvrA (AA 88-505) for bacterial 2-hybrid
HM3691	5'-CGGCCACGATGCGTCCGGCGTAGAGTTATAACCCGGCGCCTATCT -3'	amplification of *S. typhimurium* UvrA (AA 88-505) for bacterial 2-hybrid
HM3692	5'- ATGACCGCACCGGGGC -3'	amplication of *Mtb* Mfd
HM3693	5'- ACCACGCGGTGTGCGGTTCCGGTGCCCGGCGCGA -3'	amplication of *Mtb* Mfd
HM3694	5'- TCGCGCCGGGCACCGGAACCGCACACCGCGTGGT -3'	amplication of *Mtb* Mfd
HM3695	5'-CCGAAGGTATCGGCGCGCTCGACGATCAAAGTGTTGGC -3'	amplication of *Mtb* Mfd
HM3696	5'-GCCAACACTTTGATCGTCGAGCGCGCCGATACCTTCGG -3'	amplication of *Mtb* Mfd
HM3697	5'-CATTAGGCACCCCACTCGAGTCACGGTTGTCGCTCCTTAATC -3'	amplication of *Mtb* Mfd
HM3698	5'-GCCCCAGCTGGCAATTCCGGTCACGGTTGTCGCTCCTTAATC -3'	amplication of *Mtb* Mfd
HM3702	5'- CGGTACCTCGAGTCTAGAAGATCCCGATGGACCGGGTGA -3'	amplification of 3' CDS of *Mtb* mfd for construction of pHM566 (pKM342-mfd plasmid)

HM3703	5'-ACCGAGCTCGAATTAAGTACTTCCAGAACCGTTGCACGGT -3'	amplification of 3' CDS of *Mtb* mfd for construction of pHM566 (pKM342-mfd plasmid)
HM3717	5'- TAAATAAGCTAGCTTTACCTGTTCGGCGC -3'	amplification of *S. typhimurium* for clonining into pUC19
HM3718	5'- TAAATAACTCGAG TTATGCAATAGCGTTTTCTTC -3'	amplification of *S. typhimurium* for clonining into pUC19
HM3719	5'- TTATTTACTCGAGTGGGGTGCCTAATGAGT -3'	amplification of pUC19 backbone
HM3720	5'- TTATTTAGCTAGCCGGTATTTTCTCCTTACGC -3'	amplification of pUC19 backbone
HM3725	5'- CCGCCAGCCTCCCGAGTGGTCATGC -3'	site directed mutagenesis of pHM481 and pHM457 (L499R mutation)
HM3726	5'- GCATGACCACTCGGGAGGCTGGCGG -3'	site directed mutagenesis of pHM481 and pHM457 (L499R mutation)
HM3729	5'-CAACAGCGCGCCGGCCGTCGCGTATTCG -3'	site directed mutagenesis of pHM499 and pHM650 (R165A mutation)
HM3730	5'-CGAATACGCGACGGCCGGCGCGCTGTTG -3'	site directed mutagenesis of pHM499 and pHM650 (R165A mutation)
HM3908	5'- AGTTCTAGAGCGGCCGCCA -3'	amplification of hygromycin marked cassette (pHM4566) to make Mfd deletion in *Mtb*
HM4007	5'-AGTGGCCTGAAGAGACGTTTGGCGCAGCGGCCGCACGTGAATTCCTGGAAGCCTC -3'	amplification of *Mtb* Mfd (AA 334-651 domain) for bacterial 2-hybrid

HM4011	5'-CTGCGATGCAGATCTGTAAGGTAAGTTACGACCCGTCGGCCAG -3'	amplification of *Mtb* Mfd (AA 334-651 domain) for bacterial 2-hybrid
HM4064	5'-ACAAACCAGGCATCCGAGTCTCTACCGGGAAAGGTGAATGGTGTAGGCTGGAGCTGCTTC -3'	amplification of pkd13 template to make uvrA deletion in *S. typhimurium*
HM4065	5'-CCGCTTCCCGGCGAATCGGAAAGCGGCGTGAAGGTTACATATGAATATCCTCCTTAG- '3	amplification of pkd13 template to make uvrA deletion in *S. typhimurium*
HM4759	5'- TTTTGGCGGATGAGAGAAGA -3'	linearization of pMMB67EH backbone for construction of pHM649 and pHM651
HM4760	5'- ACAGCTCATTTCAGAATATTTG -3'	linearization of pMMB67EH backbone for construction of pHM649 and pHM651
HM4761	5'-GAAAATCTTCTCTCATCCGCCAAAACTTTACCTGTTCGGCGCGTT -3'	amplifcation of *S. typhimurium* mfd for construction of pHM649 and pHM651
HM4762	5'-CTGGCAAATATTCTGAAATGAGCTGTTTATGCAATAGCGTTTTCTT -3'	amplifcation of *S. typhimurium* mfd for construction of pHM649 and pHM651
HM4896	5'- ATGGTTTACTCCTATACCGAG -3'	PCR and sequencing primer for *S. typhimurium rpoB*
HM4897	5'- CTCGTCTTCCAGTTCGATG -3'	PCR and sequencing primer for *S. typhimurium rpoB*
HM4898	5'- TAGATTTCTACCAGCGCGC -3'	sequencing primer for *S. typhimurium rpoB*

HM4899	5'- TACCGCATGATGCGCCCT -3'	sequencing primer for *S. typhimurium rpoB*
HM4900	5'- TTGTCAGCGCGCAGAGTCG -3'	sequencing primer for *S. typhimurium rpoB*
HM4901	5'- CCGCTGGTGGGTACCGGTA -3'	sequencing primer for *S. typhimurium rpoB*
HM4902	5'- ACCCGGCTGGATACGGCGT -3'	sequencing primer for *S. typhimurium rpoB*
HM4903	5'- GATAAGATGGCAGGTCGTCA -3'	sequencing primer for *S. typhimurium rpoB*
HM4993	5'- GGTGGAAGAGATGGCGAAAG -3'	PCR and sequencing primer for *S. typhimurium folA*
HM4994	5'- CTTTCCTAACGGTCAACTGG -3'	PCR and sequencing primer for *S. typhimurium folA*

5.3 Strains used and constructed

B. subtilis stains used in this study were derivates of the HM1 (JH642) parent strain[206]. Transformations into *B. subtilis* HM1 were performed under standard conditions as previously described[207]. Plasmids used in this study were grown in *E. coli* DH5α. Plasmids were cloned using chemical transformations of competent *E. coli*. Plasmid purification was performed by growth of appropriate *E. coli* strain overnight at 37° C in Luria-Bertani (LB) medium supplement with the appropriate antibiotic. Plasmids were subsequently purified using the GeneJet Plasmid Miniprep Kit (Thermo). Strains for the following species were built as described: deletions in *B. subtilis* were built with transformation of marked genomic DNA into the appropriate background strain. Deletions in *S. typhimurium* were built using the λ-red recombineering [208] and all

plasmids used were transformed by electroporation. The *mfd* deletion in *P. aeruginosa* was built using the pEX18 suicide plasmid (with homology regions to *mfd*) as previously described [209], and the *mfd* deletion in *Mtb* was built using recombineering [210]. *E. coli* strains for the bacterial 2-hybrid assay were built via electroporation of the designated plasmid into the appropriate strain background. Specific details of strains and plasmid constructed used in this work, including primers used, are listed in the materials and methods chapter. All strain modifications were confirmed by PCR and sequencing.

5.4 Growth conditions

For experiments in *B. subtilis* and *S. typhimurium*, cultures were grown as described unless otherwise indicated. Cells were plated on LB supplemented with the appropriate antibiotic for isolation of single colonies. Overnight cultures from single colonies were grown at 37° C in LB at 260 RPMs and the following day cells were diluted back to OD600 0.05 and grown to exponential phase (OD600 0.3-0.5) before harvesting. For acute rifampicin ChIP experiments, cultures were grown in identical fashion until they reach OD600 0.3-0.5 and rifampicin was added at a concentration of 100µg/mL for 5 minutes before harvesting.

5.5 ChIP-seq experiments

For *B. subtilis* Mfd ChIP experiments, c-Myc mouse monoclonal antibody (clone 9E10) was used (Thermo). For S. typhimurium Mfd-GFP ChIP experiments, Green Fluorescent Protein (GFP) rabbit polyclonal antibody was used (abcam). For RpoB experiments in *B. subtilis* and *S. typhimurium*, RNA polymerase beta mouse monoclonal antibody (clone 8RB13) was used (Thermo).

ChIP experiments were performed as previously described[73,211]. Briefly, Cells were grown to exponential phase as previously described and crosslinked with 1% formaldehyde v/v. After 20 minutes at room temperature, .5M final concentration of glycine was added and cells were pelleted, washed in cold 1x PBS and pelleted again. Cells were resuspended in solution A (10 mM Tris pH 8.0, 10 mM EDTA, 50 mM NaCl, 20% sucrose) supplemented with 1 mg/ml lysozyme and 1 mM AEBSF at 37° C for 30 minutes. 2x IP buffer (100 mM Tris pH 7.0, 10 mM EDTA, 300 mM NaCl, 20% triton x-100), supplemented with 1mM AEBSF, was then added and lysates were incubated on ice for 30 minutes. Cell lysates were sonicated four times as 30% amplitude for ten seconds using a Fisher sonic dismembrator (Fisher FB120). Lysates were centrifuged at 8000 RPMs for 15 minutes at 4° C. The supernatant was transferred into new microfuge tubes.

ChIP lysates were split into a total DNA input control (40µl of lysate) and immunoprecipitation (IP) (1mL of lysate). For Mfd ChIP experiments 12µl anti-c-Myc antibody and 1µl of anti-GFP were added to the IP samples and 2µl of anti-RpoB antibody was added for RpoB ChIPs. IP lysates were rotated overnight at 4° C. The following day, 30µl Protein A sepharose beads (GE) were added to the IP samples and rotated for one hour at room temperature. Beads were then pelleted with centrifugation at 2000 RPMs for one minute. Supernatant was decanted and beads were subsequently washed six times with 1x IP buffer and one time with 1x TE pH 8.0. Beads were then pelleted and resuspended in 100µl of elution buffer (50mM Tris pH 8.0, 10mM EDTA), and 1% SDS and incubated at 65° C for 10 minutes. Beads were pelleted by centrifugation and supernatant was transferred to a new microfuge tube. A second round of elution was performed by resuspension of beads in 150µl of elution buffer II (10mM Tris pH 8.0, 1 mM EDTA, 0.67% SDS). Beads were pelleted and supernatant was transferred to microfuge

tube containing eluate from the first elution. IP samples were then incubated overnight at 65° C. The following day, proteinase K was added at a final concentration 0.4 mg/mL and samples were incubated for two hours at 37° C. Purification was performed by using the GeneJet PCR Purification Kit (Thermo).

Library preparation for ChIP-seq was performed using the Nextera XT DNA Library Prep Kit (Illumina) according to manufacturer's instructions.

5.6 RNA-seq experiments

B. subtilis cultures were grown to exponential growth as previously described and harvested by addition of 1:1 volume 100% cold methanol and centrifugation at 5000RPMs for 5minutes. Cell pellets were subsequently lysed in TE and lysozyme (20mg/mL) and purified using the GeneJet RNA Purification Kit (Thermo). Library preparation for RNA-seq was performed using the Scriptseq Complete Kit (Bacteria) from Illumina, according to manufacturer's instructions.

5.7 Whole-genome sequencing analysis

ChIP-seq and RNA-seq samples were sequenced using the Illumina Nextseq 500/550 Sequencing system at the University of Washington Northwest Genomics Center and the VANTAGE Sequencing Core at Vanderbilt University. After sequencing, sample reads from *B. subtilis* were mapped to *B. subtilis* 168 genome (accession number: NC_000964.3) and *S. typhimurium* reads were mapped to *S. typhimurium* genome LT2 (accession number: AE006468.2) using Bowtie2[212]. For data visualization, SAMtools was used to process SAM files[213] to produce wiggle plots[214]. Wiggle files from all ChIP samples were normalized to input samples (total input DNA subtracted from the ChIP signal). For quantification of ChIP-seq and

RNA-seq samples, BAM files were processed by the featureCounts program to determine read counts per gene[215]. To determine differential RNA-seq expression and differential ChIP-seq binding, read counts were analyzed by DEseq2 software[216]. In order to determine correlations between RpoB ChIP binding and Mfd ChIP binding (in both *B. subtilis* and *S. typhimurium*), read counts generated by featureCounts were divided by the total number of sequencing reads per sample. ChIP samples were then divided by input samples and $\log_2$ normalized.

5.8 Quantitative RT-PCR assays

For quantification of RNA using qRT-PCR, RNA was extracted as described in the RNA-seq experiments section. Subsequently, 1µg of total RNA was treated with DNaseI (Thermo) for one hour at 37° C. DNase denaturation was performed with addition of 10mM EDTA and incubation at 65° C for 10 minutes. cDNA generation was performed using the iScript Supermix (BioRad), according to the manufacturer's instructions. Quantitative PCR (qPCR) was performed using the Sso Advanced Universal SYBR Green Supermix (BioRad), according to manufacturer's instructions. For normalization of qRT-PCR, primers to *B. subtilis* rRNA was used.

5.9 Toxin cell survival assays

For chronic survival assays, strains were struck out on LB agar plates and *B. subtilis* cultures were grown in 2mL LB until they reached an OD600 of 0.5-1.0. All cultures were normalized to OD600 0.3 and serial dilutions were performed in 1x Spizizen's salts. 5µl of cells were plated on control plates containing LB agar only and LB agar plates containing the designated concentration of IPTG (see figure legends for concentrations). Plates were grown at 30° C overnight and CFUs were enumerated the following day.

For acute survival assays, cultures were grown in 2ml until they reached an OD600 of 0.5-1.0 and then diluted back to OD600 0.05. Either was 1mM or .1 mM IPTG was then added and cells were grown for approximately 60 minutes (OD600~.3). Cells were subsequently washed two times with 1x Spizizen's salts to remove residual IPTG and were serially diluted. 5µl of cells were plated on LB agar and plates were grown at 30° C overnight for CFU enumeration. For both chronic and acute survival assays, images were taken using the BioRad Gel Doc XR+ Molecular Imager.

5.10 Western blot assays

Exponentially growing cultures were centrifuged, resuspended in Tris/Salt buffer (50 mM Tris-HCl pH 8, 300mM NaCl), and pelleted. Cell lysis buffer (10mM Tris-HCl pH7, 10mM EDTA, .1mM AEBSF, .1mg/ml lysozyme) was added and samples were incubated at 37° C for 15 minutes. SDS loading buffer was added to samples and 20µl was loaded onto Mini-PROTEAN TGX Precast Gels (BioRad) and run in Tris/SDS/Glycine running buffer in a Mini-PROTEAN Electrophoresis Cell (BioRad) at 200V for 40 minutes. Transfer was performed using the Trans-Blot Turbo Transfer System (BioRad). Anti-c-Myc antibody (1:5000 dilution) and anti-GFP antibody (1:7500 dilution) was added and blots were incubated overnight at 4° C. Anti-mouse and anti-rabbit secondary antibodies (Li-Cor) (1:15000 dilution) were added to Mfd-myc and Mfd-GFP western, respectively, and blot was imaged using the Odyssey CLx imaging system (Li-Cor).

5.11 Toxin-antitoxin mutation rate analysis and sequencing of revertants

Luria-Delbrück mutation rate assays were performed as previously described[45]. B. subtilis were grown on LB plates overnight and cultures from single colonies and subsequently grown in LB media at 37° C at 260 RPMs to exponential phase growth (OD600= .5). Cells were diluted back to OD600= 0.0005 and dispended into 2mL parallel cultures containing 2mL LB and grown in the same conditions until OD600=0.5. To identify mutants that were resistant to toxin overexpression, 100µl of each 2ml culture was plated on LB plates containing 1mM IPTG. Cells were serially diluted and plated on LB for CFU enumeration. Colonies were quantified after growth overnight at 37° C for IPTG plates and 30° C for LB plates. Mutation rates were calculated using the Ma-Sandri-Sarkar Maximum Likelihood method[217].

For sanger sequencing of revertants, colonies were picked at random and grown in LB supplement with .5mM IPTG at 37° C until cultures reached OD600=0.5. Genomic DNA purification was performed using the GeneJet DNA Purification Kit, per manufacturer's instructions. PCR amplification was performed using Phusion DNA polymerase (Thermo) and purified using the GeneJet PCR Purification Kit (Thermo) and subsequently sequenced (Genewiz).

5.12 Luria-Delbruck fluctuation analysis

For *B. subtilis*, cultures were grown from single colonies at 37°C with aeration in LB media (10g Tryptone, 5g yeast extract and 5g NaCl per liter). Exponential phase cultures (OD = 0.3) were diluted back to OD = 0.0005 in parallel cultures containing LB, and plated following 4.5 hours of growth at 37°C with aeration. Cultures were plated on 50 µg/mL rifampicin to quantify the number

of mutants and serially diluted and plated on LB to quantify total viable cells. Colonies were quantified after overnight incubation at 37°C (for rifampicin plates) and 30°C (for LB plates).

S. typhimurium and *P. aeruginosa* mutation rates were measured by growing overnight cultures from single colonies and subsequently back diluting parallel cultures to an OD_{600} = 0.0005 in LB. Cultures were grown to an OD_{600} = 0.8-1.0 (OD_{600} = 1.0 for *P. aeruginosa*) at 37°C with aeration. Cultures were plated as described for *B. subtilis*. For mutation rate analysis of WT-pUC19, Δ*mfd*-pUC19, ST19 *mfd* complementation, *Mtb mfd* complementation, and the ST19 *mfd* point mutant (L499R and R165A) strains of *S. typhimurium*, overnight cultures were grown to saturation in 50 μg/mL carbenicillin to maintain plasmid selection. Cultures were back diluted to an OD_{600} = 0.0005, grown in LB only to OD_{600} = 0.8-1.0 and plated as previously described.

For *Mtb*, experiments were performed as previously described [44]. Briefly, cultures were grown in 7H9 mycobacterial agar + ADC to saturation. Multiple, independent cultures were back diluted to final OD_{600}= 0.0001 and grown at 37°C to OD_{600}= 0.8-1.2. Cells were plated on 7H10 mycobacterial agar + OADC and 2μg/mL rifampicin, 5μg/mL ethambutol or 1.5μg/mL of ciprofloxacin to quantify resistant mutants and on 7H10 + OADC for CFU enumeration. Plates were incubated at 37°C for approximately 10 days for CFU enumeration and 25-30 days for antibiotic plates. Mutation rates for all species were calculated using the Ma-Sandri-Sarkar Maximum Likelihood method [217].

5.13 Mutagenesis measurements post epithelial cell infection

Colorectal adenocarcinoma cells line CACO-2 were cultured in DMEM medium with 20% heat-inactivated FBS at 37°C in 5% CO_2. Approximately 10^6 CACO-2 cells were plated overnight in 6-well plates at 37°C in 5% CO_2 for infection. A single *S. typhimurium* colony was picked and grown

overnight at 37°C in LB, diluted back to an OD_{600} = 0.05 the following day and grown at 37°C in LB until cultures reached OD_{600} = 0.5. Cells were washed 2x with 1X PBS resuspended in DMEM +20% FBS and inoculated at 100:1 multiplicity of infection with CACO-2 cells at 37°C in 5% CO_2 for one hour. Cells were then washed 2x with 1X PBS and DMEM + 20% FBS + 50 µg/mL gentamicin was added to plates to kill extracellular bacteria. After 6 hours of infection, cells were washed in 1X PBS and lysed in 1X PBS + 0.1% Triton X-100. Cells were plated on M9 minimal + 0.4% glycerol agar for CFU enumeration and M9 minimal + 0.4% glycerol agar containing 100µg/mL 5-fluorocytosine (5FC) and grown at 37°C to determine mutation frequency. Mutation frequency was determined by taking the ratio of 5FC colonies to the viable cell count for each sample. For experiments measuring cell viability over multiple time points, *S. typhimurium* and CACO-2 cells were grown as described and bacterial cells were harvested for CFU enumeration at defined time points.

5.14 Antibiotic evolution assays

Evolution experiments were performed for the indicated strains. For *S. typhimurium*, overnight cultures, started from a single colony, were back diluted to OD_{600} = 0.005 and used to inoculate a 96-well plate. Cells were grown for either 12 or 24 hours with agitation, at 37°C, in LB with a gradient of concentrations of the indicated antibiotic to select for resistance. ODs were subsequently measured in an Epoch/2 microplate spectrophotometer (Bio-Tek). Cultures that grew (defined by at least 50% growth relative to LB only) at the highest concentration of antibiotic were passaged into fresh LB +antibiotic in a subsequent plate. A total of 5-8 serial passages were performed depending on the antibiotic used. Evolution experiments with WT-pMMB67EH, Δ*mfd* –pMMB67EH, complementation and point mutant (L499R and R165A) strains of *S. typhimurium*

were grown identical to WT and Δ*mfd* strains except with the addition of 50 μg/mL carbenicillin to maintain selection of episomes. For *B. subtilis*, cultures were started from a single colony were grown for 4-5 hours until they reached OD_{600} = 1.0. Cultures were back diluted to OD_{600} = 0.005, inoculated into a 96-well plate and grown for 12 hours at 37°C in LB in an Epoch/2 microplate spectrophotometer (Bio-Tek) for 9 serial passages, with a gradient of concentrations of rifampicin to select for resistance. For *Mtb*, saturated cultures were back diluted to OD_{600} = 0.05, inoculated into a 96-well plate and grown in 7H9 +ADC in a 37°C incubator without aeration. Strains were serially passaged when the density of the no antibiotic control wells reached approximately OD_{600} =1.5-2.0 (approximately 15-20 days). Cultures that grew (defined by at least 50% growth relative to 7H9+ADC) at the highest concentration of rifampicin were passaged into a fresh 7H9 + ADC+ rifampicin in a subsequent plate, and a total of 6 serial passages were performed. For all species, antibiotics were diluted 2-fold down each given row in a 96 well plate.

5.15 Sequencing of antibiotic evolution assays

Genomic DNA was harvested from evolved strains of *S. typhimurium* and purified using either the MasterPure complete DNA and RNA Purification Kit (Epicentre) or GeneJet Genomic DNA Purification Kit (ThermoFisher) in accordance with manufacturer instructions. For WGS experiments, gDNA samples were processed for sequencing using the Nextera XT DNA Library Preparation Kit (Illumina). Paired-end libraries were sequenced on an Illumina NextSeq sequencing platform yielding an average of 40X coverage sequencing depth. The resulting FASTQ reads were trimmed for quality using the FASTX quality filter such that 95% of bases were required to have a Phred score of 30 or higher (*Available at http://hannonlab.cshl.edu/fastx_toolkit/index.html*). SNPs against the *S. typhimurium* ST19

genome (available from the Prokaryotic Genome Analysis Tool http://tools.uwgenomics.org/pgat/), a derivative of *S. enterica* Typhimurium LT2 (Genbank Accession NC_003197.2) were then identified using BreSeq [218]. For Sanger sequencing, amplification of *rpoB* and *folA* loci was performed using Phusion DNA polymerase (ThermoFisher). PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) and sequencing was subsequently performed to identify mutations.

5.16 Bacterial 2-hybrid assays

Bacterial 2-hybrid assays were performed as previously described [219]. Briefly, the genes of interest were fused to the Lambda repressor (cI) and the N-terminal domain of *E. coli* RNA polymerase's alpha subunit (α-NTD) using the plasmids pACλCI and pBRα, respectively. These fusion constructs were transformed into *E. coli* containing the Lambda operator sequence inserted upstream of a luciferase reporter (NanoLuc, Promega) using an F' episome. For expression of fusion constructs, cells were grown overnight in LB + 20μM IPTG at 30°C and diluted 1:100 into fresh LB + 20μM IPTG at 30°C the next morning and were grown until OD_{600}= 2. For relative light unit measurements, Nano-glo substrate (Promega) was added to cultures according to the manufacturer's instructions and luminescence was measured in a SpectraMax M3 96-well plate reader.

5.17 DNA damage survival assays

For both *S. typhimurium* and *B. subtilis*, cultures were started from single colonies and harvested at exponential growth (OD_{600} = 0.3-0.6). To determine 4-Nitro-Quinolone Oxide (4-NQO) survival, cell dilutions were spotted onto LB agar plates (for CFU enumeration) and LB agar plates

containing either 0.2 μM (*B. subtilis*) or 4 μM (*S. typhimurium)* 4-NQO. To determine UV sensitivity, cells were spotted onto LB agar plates and exposed to the indicated intensity of UV light using a Mineralight XX 15V UV light source (UVP). Surviving colonies were enumerated after overnight incubation at 30°C.

5.18 Bone Marrow-Derived Macrophage (BMM) infections

BMMs were derived from BALB/c mice as previously described [220]. BMMs were cultured in RPMI media with 10% heat-inactivated FBS at 37°C in 5% CO_2. Approximately 10^6 BMMs were plated overnight in 24-well plates at 37°C in 5% CO_2. For infections, a single colony of *S. typhimurium* was picked and grown overnight at 37°C in LB, diluted back to an OD_{600} = 0.05 the following day and grown at 37°C in LB until cultures reached OD_{600} = 0.5. Bacteria were then washed 2X with 1X PBS and resuspended in RPMI +10% FBS and inoculated at 10:1 multiplicity of infection with BMMs at 37°C in 5% CO_2 for 30 minutes. Plates were then washed 2X with 1X PBS and RPMI +10% FBS + 50 μg/mL gentamicin was added for killing of extracellular bacteria. Infected macrophages were lysed at indicated time points with 1X PBS + 0.1% Triton X-100 and plated on LB for CFU enumeration.

Chapter 6

References

1. Darwin, C. *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. (1859).

2. Cooper, G. M. Heredity, Genes, and DNA. (2000).

3. Luria, S. & Delbrück, M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28,** 491–511 (1943).

4. Foster, P. L. Adaptive mutation: the uses of adversity. *Annu. Rev. Microbiol.* **47,** 467–504 (1993).

5. Foster, P. L. Adaptive mutation: implications for evolution. *BioEssays* **22,** 1067–1074 (2000).

6. Hall, B. G. On the specificity of adaptive mutations. *Genetics* **145,** 39–44 (1997).

7. Hall, B. G. Adaptive mutations in Escherichia coli as a model for the multiple mutational origins of tumors. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 5669–73 (1995).

8. Cairns, J., Overbaugh, J. & Miller, S. The origin of mutants. *Nature* **335,** 142–145 (1988).

9. Cairns, J. & Fostert, P. L. Adaptive Reversion of a Frameshift Mutation. *Genetics* **701,** 695–701 (1991).

10. Rosenberg, S. M. Evolving responsively: adaptive mutation. *Nat. Rev. Genet.* **2,** 504–515 (2001).

11. Roth, J. R., Kugelberg, E., Reams, A. B., Kofoid, E. & Andersson, D. I. Origin of Mutations Under Selection: The Adaptive Mutation Controversy. *Annu. Rev. Microbiol.* **60,** 477–501 (2006).

12. Lodish H, Berk A, Zipursky SL, et al. *Molecular Cell Biology. 4th edition*. (2000).

13. Prescott, D. M. & Kuempel, P. L. Bidirectional Replication of the Chromosome in Escherichia coli. *Proc. Natl. Acad. Sci.* **69,** 2842–2845 (1972).

14. Mchenrys, C. & Kornberg, A. *DNA Polymerase III Holoenzyme of Escherichia COB PURIFICATION AND RESOLUTION INTO SUBUNITS*. THE JOURNAL OF BIOLOGICAL CHEMI~TRV* **252,** (1977).

15. Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* **461,** 1243–7 (2009).

16. Ganai, R. A. & Johansson, E. DNA Replication-A Matter of Fidelity. *Mol. Cell* **62,** 745–755 (2016).

17. Fijalkowska, I. J., Schaaper, R. M. & Jonczyk, P. DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair. *FEMS Microbiol. Rev.* **36,** 1105–21 (2012).

18. Scheuermann, R., Tam, S., Burgers, P. M., Lu, C. & Echols, H. Identification of the epsilon-subunit of Escherichia coli DNA polymerase III holoenzyme as the dnaQ gene product: a fidelity subunit for DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 7085–9 (1983).

19. Rock, J. M. *et al.* DNA replication fidelity in Mycobacterium tuberculosis is mediated by an ancestral prokaryotic proofreader. *Nat. Genet.* **47,** 677–681 (2015).

20. Marti, T. M., Kunz, C. & Fleck, O. DNA mismatch repair and mutation avoidance pathways. *J. Cell. Physiol.* **191,** 28–41 (2002).

21. Putnam, C. D. Evolution of the methyl directed mismatch repair system in Escherichia coli. *DNA Repair (Amst).* **38,** 32–41 (2016).

22. Cooper, D. L., Lahue, R. S. & Modrich, P. Methyl-directed mismatch repair is bidirectional. *J. Biol. Chem.* **268,** 11823–9 (1993).

23. Bronner, C. E. *et al.* Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer. *Nature* **368,** 258–261 (1994).

24. Michaels, M. L., Cruz, C., Grollman, A. P. & Miller, J. H. Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 7022–5 (1992).

25. Jena, N. R. DNA damage by reactive species: Mechanisms, mutation and repair. *J. Biosci.* **37,** 503–17 (2012).

26. Nakabeppu, Y. Cellular levels of 8-oxoguanine in either DNA or the nucleotide pool play pivotal roles in carcinogenesis and survival of cancer cells. *Int. J. Mol. Sci.* **15,** 12543–57 (2014).

27. Leo, M. & Miller, J. H. *The GO System Protects Organisms from the Mutagenic Effect of the Spontaneous Lesion 8-Hydroxyguanine (7,8-Dihydro-8-Oxoguanine)*. *JOURNAL OF BACrERIOLOGY* **174,** (1992).

28. Kohanski, M. A., Depristo, M. A. & Collins, J. J. Sublethal Antibiotic Treatment Leads to Multidrug Resistance via Radical-Induced Mutagenesis. *Mol. Cell* **37,** 311–320 (2010).

29. Radman, M. SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis. *Basic Life Sci.* **5A,** 355–67 (1975).

30. Simmons, L. A., Foti, J. J., Cohen, S. E. & Walker, G. C. The SOS Regulatory Network. *EcoSal Plus* **2008,** (2008).

31. Smith, B. T. & Walker, G. C. Mutagenesis and more: umuDC and the Escherichia coli SOS response. *Genetics* **148,** 1599–610 (1998).

32. Reuven, N. B., Tomer, G. & Livneh, Z. The mutagenesis proteins UmuD' and UmuC prevent lethal frameshifts while increasing base substitution mutations. *Mol. Cell* **2,** 191–9 (1998).

33. Reuven, N. B., Arad, G., Maor-Shoshani, A. & Livneh, Z. The mutagenesis protein UmuC is a DNA polymerase activated by UmuD', RecA, and SSB and is specialized for translesion replication. *J. Biol. Chem.* **274,** 31763–6 (1999).

34. Henrikus, S. S. *et al.* DNA polymerase IV primarily operates outside of DNA replication forks in Escherichia coli. *PLoS Genet.* **14,** e1007161 (2018).

35. Boshoff, H. I. M., Reed, M. B., Iii, C. E. B., Mizrahi, V. & Drive, P. DnaE2 Polymerase Contributes to In Vivo Survival and the Emergence of Drug Resistance in Mycobacterium tuberculosis. *Cell* **113,** 183–193 (2003).

36. Aranda, J. *et al.* Role of Acinetobacter baumannii UmuD Homologs in Antibiotic Resistance Acquired through DNA Damage-Induced Mutagenesis. *Antimicrob. Agents Chemother.* **58,** 1771–1773 (2014).

37. Sanders, L. H., Rockel, A., Lu, H., Wozniak, D. J. & Sutton, M. D. Role of Pseudomonas aeruginosa dinB-encoded DNA polymerase IV in mutagenesis. *J. Bacteriol.* **188,** 8573–85 (2006).

38. Foster, P. L. Stress-induced mutagenesis in bacteria. *Crit. Rev. Biochem. Mol. Biol.* **42,** 373–97 (2007).

39. Layton, J. C. & Foster, P. L. Error-prone DNA polymerase IV is controlled by the stress-response sigma factor, RpoS, in Escherichia coli. *Mol. Microbiol.* **50,** 549–61 (2003).

40. Feng, G., Tsui, H. C. & Winkler, M. E. Depletion of the cellular amounts of the MutS and MutH methyl-directed mismatch repair proteins in stationary-phase Escherichia coli K-12

cells. *J. Bacteriol.* **178,** 2388–96 (1996).

41. Tsui, H. C., Feng, G. & Winkler, M. E. Negative regulation of mutS and mutH repair gene expression by the Hfq and RpoS global regulators of Escherichia coli K-12. *J. Bacteriol.* **179,** 7476–87 (1997).

42. Sung, H.-M. & Yasbin, R. E. Adaptive, or stationary-phase, mutagenesis, a component of bacterial differentiation in Bacillus subtilis. *J. Bacteriol.* **184,** 5641–53 (2002).

43. Ross, C. *et al.* Novel Role of mfd : Effects on Stationary-Phase Mutagenesis in Bacillus subtilis. *J. Bacteriol.* **188,** 7512–7520 (2006).

44. Ford, C. B. *et al.* Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat. Genet.* **45,** 784–90 (2013).

45. Ragheb, M. N. *et al.* Inhibiting the Evolution of Antibiotic Resistance. *Mol. Cell* **73,** 157–165.e5 (2019).

46. Oliver, A. *et al.* High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science (80-.).* **288,** 1251–4 (2000).

47. Baym, M. *et al.* Spatiotemporal microbial evolution on antibiotic landscapes. *Science (80-.).* **353,** 1147–1152 (2016).

48. Fitch, W. M. Evidence suggesting a non-random character to nucleotide replacements in naturally occurring mutations. *J. Mol. Biol.* **26,** 499–507 (1967).

49. Rosenberg, M. S., Subramanian, S. & Kumar, S. Patterns of Transitional Mutation Biases Within and Among Mammalian Genomes. *Mol. Biol. Evol.* **20,** 988–993 (2003).

50. Topal, M. D. & Fresco, J. R. Complementary base pairing and the origin of substitution

mutations. *Nature* **263,** 285–289 (1976).

51. Griffiths, A. J. F. *Modern genetic analysis*. (W.H. Freeman, 1999).

52. Jee, J. *et al.* Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. *Nature* **534,** 693–696 (2016).

53. Wang, G. & Vasquez, K. M. Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst).* **19,** 143–151 (2014).

54. Shah, K. A. & Mirkin, S. M. The hidden side of unstable DNA repeats: Mutagenesis at a distance. *DNA Repair (Amst).* **32,** 106–112 (2015).

55. Bidichandani, S. I. *et al.* Somatic sequence variation at the Friedreich ataxia locus includes complete contraction of the expanded GAA triplet repeat, significant length variation in serially passaged lymphoblasts and enhanced mutagenesis in the flanking sequence. *Hum. Mol. Genet.* **8,** 2425–36 (1999).

56. Shishkin, A. A. *et al.* Large-Scale Expansions of Friedreich's Ataxia GAA Repeats in Yeast. *Mol. Cell* **35,** 82–92 (2009).

57. Aksenova, A. Y. *et al.* Genome rearrangements caused by interstitial telomeric sequences in yeast. *Proc. Natl. Acad. Sci.* **110,** 19866–19871 (2013).

58. Lin, Y. & Wilson, J. H. Transcription-Induced CAG Repeat Contraction in Human Cells Is Mediated in Part by Transcription-Coupled Nucleotide Excision Repair. *Mol. Cell. Biol.* **27,** 6209–6217 (2007).

59. Tang, W. *et al.* Friedreich's Ataxia (GAA)n•(TTC)n Repeats Strongly Stimulate Mitotic Crossovers in Saccharomyces cerevisae. *PLoS Genet.* **7,** e1001270 (2011).

60. Tang, W., Dominska, M., Gawel, M., Greenwell, P. W. & Petes, T. D. Genomic deletions

and point mutations induced in Saccharomyces cerevisiae by the trinucleotide repeats (GAA·TTC) associated with Friedreich's ataxia. *DNA Repair (Amst).* **12,** 10–17 (2013).

61. Shah, K. A. *et al.* Role of DNA Polymerases in Repeat-Mediated Genome Instability. *Cell Rep.* **2,** 1088–1095 (2012).

62. Petermann, E. & Helleday, T. Pathways of mammalian replication fork restart. *Nat. Rev. Mol. Cell Biol.* **11,** 683–687 (2010).

63. Rodgers, K. & McVey, M. Error-Prone Repair of DNA Double-Strand Breaks. *J. Cell. Physiol.* **231,** 15–24 (2016).

64. Deem, A. *et al.* Break-Induced Replication Is Highly Inaccurate. *PLoS Biol.* **9,** e1000594 (2011).

65. Kramara, J., Osia, B. & Malkova, A. Break-Induced Replication: The Where, The Why, and The How. *Trends Genet.* **34,** 518–531 (2018).

66. Saini, N. *et al.* Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502,** 389–392 (2013).

67. Xie, K. T. *et al.* DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science (80-.).* **363,** 81–84 (2019).

68. Schroeder, J. W., Hirst, W. G., Szewczyk, G. A. & Simmons, L. A. The Effect of Local Sequence Context on Mutational Bias of Genes Encoded on the Leading and Lagging Strands. *Curr. Biol.* **26,** 692–697 (2016).

69. Merrikh, C. N., Weiss, E. & Merrikh, H. The Accelerated Evolution of Lagging Strand Genes Is Independent of Sequence Context. *Genome Biol. Evol.* **8,** evw274 (2016).

70. Jinks-Robertson, S. & Bhagwat, A. S. Transcription-Associated Mutagenesis. *Annu. Rev.*

Genet. **48,** 341–359 (2014).

71. Thomas, M., White, R. L. & Davis, R. W. Hybridization of RNA to double-stranded DNA: Formation of R-loops. *Proc. Natl. Acad. Sci.* **73,** 2294–2298 (1976).

72. Kim, N. & Jinks-Robertson, S. Transcription as a source of genome instability. *Nat. Rev. Genet.* **13,** 204–214 (2012).

73. Lang, K. S. *et al.* Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival. *Cell* **170,** 787–790.e18 (2017).

74. Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* **170,** 774–786.e19 (2017).

75. Aguilera, A. & García-muse, T. R Loops: From Transcription Byproducts to Threats to Genome Stability. *Mol. Cell* **46,** 115–124 (2012).

76. Wimberly, H. *et al.* R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. *Nat. Commun.* **4,** 2115 (2013).

77. Lang, K. S. & Merrikh, H. The Clash of Macromolecular Titans: Replication-Transcription Conflicts in Bacteria. *Annu. Rev. Microbiol.* **72,** 71–88 (2018).

78. Merrikh, H., Zhang, Y., Grossman, A. D. & Wang, J. D. Replication-transcription conflicts in bacteria. *Nat. Rev. Microbiol.* **10,** 449–58 (2012).

79. Hamperl, S. & Cimprich, K. A. Conflict Resolution in the Genome: How Transcription and Replication Make It Work. *Cell* **167,** 1455–1467 (2016).

80. García-muse, T. & Aguilera, A. Transcription–replication conflicts: how they occur and how they are resolved. *Nat. Rev. Mol. Cell Biol.* **17,** 553–563 (2016).

81. Mangiameli, S. M., Merrikh, C. N., Wiggins, P. A. & Merrikh, H. Transcription leads to pervasive replisome instability in bacteria. *Elife* 1–27 (2017). doi:10.7554/eLife.19848

82. De Septenville, A. L., Duigou, S., Boubakri, H. & Michel, B. Replication Fork Reversal after Replication–Transcription Collision. *PLoS Genet.* **8,** e1002622 (2012).

83. Million-Weaver, S., Samadpour, A. N. & Merrikh, H. Replication restart after replication-transcription conflicts requires RecA in Bacillus subtilis. *J. Bacteriol.* **197,** 2374–2382 (2015).

84. Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E. & Merrikh, H. Accelerated gene evolution through replication-transcription conflicts. *Nature* **495,** 512–5 (2013).

85. Million-Weaver, S. *et al.* An underlying mechanism for the increased mutagenesis of lagging-strand genes in Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E1096-105 (2015).

86. Sankar, T. S., Wastuwidyaningtyas, B. D., Dong, Y., Lewis, S. A. & Wang, J. D. The nature of mutations induced by replication–transcription collisions. *Nature* **535,** 178–181 (2016).

87. Robleto, E. A. Transcriptional De-Repression and Mfd Are Mutagenic in Stressed Bacillus subtilis Cells. 45–58 (2012). doi:10.1159/000332751

88. Hyug, G. *et al.* The Helicobacter pylori Mfd protein is important for antibiotic resistance and DNA repair. *Diagn. Microbiol. Infect. Dis.* **65,** 454–456 (2009).

89. Han, J., Sahin, O., Barton, Y. & Zhang, Q. Key Role of Mfd in the Development of Fluoroquinolone Resistance in Campylobacter jejuni. *Plos Pathog.* **4,** (2008).

90. Baharoglu, Z., Babosan, A. & Mazel, D. Identification of genes involved in low aminoglycoside-induced SOS response in Vibrio cholerae : a role for transcription stalling

and Mfd helicase. *Nucleic Acids Res.* **42,** 2366–2379 (2014).

91. Merrikh, H. Spatial and Temporal Control of Evolution through Replication–Transcription Conflicts. *Trends Microbiol.* **xx,** 1–7 (2017).

92. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* **9,** 958–70 (2008).

93. Selby, C. P. & Sancar, A. Transcription-repair coupling and mutation frequency decline. *Microbiol. Rev.* **175,** 7509–7514 (1993).

94. Mellon, I., Spivak, G. & Hanawalt, P. C. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* **51,** 241–249 (1987).

95. Mellon, I. & Hanawalt, P. C. Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature* **342,** 95–98 (1989).

96. Byrd, A. K. & Raney, K. D. Superfamily 2 helicases. *Front. Biosci. (Landmark Ed.* **17,** 2070–88 (2012).

97. Deaconescu, A. M. *et al.* Structural Basis for Bacterial Transcription-Coupled DNA Repair. *Cell* 507–520 (2006). doi:10.1016/j.cell.2005.11.045

98. Fan, J., Leroux-Coyau, M., Savery, N. J. & Strick, T. R. Reconstruction of bacterial transcription-coupled repair at single-molecule resolution. *Nature* **536,** 234–7 (2016).

99. Graves, E. T. *et al.* A dynamic DNA-repair complex observed by correlative single-molecule nanomanipulation and fluorescence. *Nat. Struct. Mol. Biol.* **22,** (2015).

100. Savery, N. J. The molecular mechanism of transcription-coupled DNA repair. *Trends Microbiol.* **15,** (2007).

101. Le, T. T. *et al.* Mfd Dynamically Regulates Transcription via a Release and Catch-Up Mechanism. *Cell* **173,** 1823 (2018).

102. Sims, R. J., Belotserkovskaya, R. & Reinberg, D. Elongation by RNA polymerase II: The short and long of it. *Genes Dev.* **18,** 2437–2468 (2004).

103. Hanawalt, P. C. DNA repair. The bases for Cockayne syndrome. *Nature* **405,** 415–416 (2000).

104. Cohen, S. E. *et al.* Roles for the transcription elongation factor NusA in both DNA repair and damage tolerance pathways in Escherichia coli. *Proc. Natl. Acad. Sci.* **107,** 15517–15522 (2010).

105. Epshtein, V. *et al.* UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **505,** 372–377 (2014).

106. Kamarthapu, V. *et al.* ppGpp couples transcription to DNA repair in E.coli. *Science (80-.).* (2016). doi:10.1126/science.aad6945

107. Deaconescu, A. M., Artsimovitch, I. & Grigorieff, N. Interplay of DNA repair with transcription : from structures to mechanisms. *Trends Biochem. Sci.* **37,** 543–552 (2012).

108. Zalieckas, J. M., Wray, L. V, Ferson, A. E. & Fisher, S. H. Transcription–repair coupling factor is involved in carbon catabolite repression of the Bacillus subtilis hut and gnt operons. *Mol. Microbiol.* **27,** 1031–1038 (1998).

109. Belitsky, B. R. & Sonenshein, A. L. Roadblock repression of transcription by Bacillus subtilis CodY. *J. Mol. Biol.* **411,** 729–743 (2012).

110. Park, J., Marr, M. T. & Roberts, J. W. E. coli Transcription Repair Coupling Factor (Mfd Protein) Rescues Arrested Complexes by Promoting Forward Translocation. *Cell* **109,** 757–767 (2002).

111. Washburn, R. S., Wang, Y. & Gottesman, M. E. Role of E .coli Transcription-Repair Coupling Factor Mfd in Nun-mediated Transcription Termination. *J. Mol. Biol.* **2836,** 655–662 (2003).

112. Ho, H. N., van Oijen, A. M. & Ghodke, H. The transcription-repair coupling factor Mfd associates with RNA polymerase in the absence of exogenous damage. *Nat. Commun.* **9,** 1–12 (2018).

113. Selby, C. P. & Sancar, A. Cockayne syndrome group B protein enhances elongation by RNA polymerase II. *Proc. Natl. Acad. Sci.* **94,** 11205–11209 (1997).

114. Tantin, D., Kansal, A. & Carey, M. Recruitment of the Putative Transcription-Repair Coupling Factor CSB / ERCC6 to RNA Polymerase II Elongation Complexes. *Mol. Cell. Biol.* **17,** 6803–6814 (1997).

115. Trautinger, B. W., Jaktaji, R. P., Rusakova, E. & Lloyd, R. G. RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Mol. Cell* **19,** 247–258 (2005).

116. Pomerantz, R. T. & Donnell, M. O. Direct Restart of a Replication Fork Stalled by a Head-On RNA Polymerase. *Science (80-.).* **327,** 590–592 (2010).

117. Pomerantz, R. T. & O'Donnell, M. The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature* **456,** 762–766 (2008).

118. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA polymerase backtracking to genome instability in E. coli. *Cell* **146,** 533–543 (2011).

119. Boubakri, H., de Setptenville, A. L., Viguera, E. & Michel, B. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J.* **29,** 145–157 (2010).

120. Ray-Soni, A., Bellecourt, M. J. & Landick, R. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.* **85,** 319–347 (2016).

121. Park, J.-S. & Roberts, J. W. Role of DNA bubble rewinding in enzymatic transcription termination. *Proc. Natl. Acad. Sci.* **103,** 4870–4875 (2006).

122. Richardson, J. P. Rho-dependent termination and ATPases in transcript termination. *Biochim. Biophys. Acta* **1577,** 251–260 (2002).

123. Morgan, W. D., Bear, D. G. & von Hippel, P. H. Rho-dependent termination of transcription. II. Kinetics of mRNA elongation during transcription from the bacteriophage lambda PR promoter. *J. Biol. Chem.* **258,** 9565–74 (1983).

124. Kriner, M. A., Sevostyanova, A. & Groisman, E. A. Learning from the Leaders: Gene Regulation by the Transcription Termination Factor Rho. *Trends Biochem. Sci.* **41,** 690–699 (2016).

125. O'Neill, J. Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. 4–16 (2014).

126. Ventola, C. L. The Antibiotic Resistance Crisis Part 1 : Causes and Threats. *P&T* **40,** 277–283 (2015).

127. Eduardo, P., Da, A. & Palomino, J. C. Molecular basis and mechanisms of drug resistance in Mycobacterium tuberculosis : classical and new drugs. *J. Antimicrob. Chemother.* **66,** 1417–1430 (2011).

128. Goldstein, B. P. Resistance to rifampicin: a review. *J. Antibiot. (Tokyo).* **67,** 625–630 (2014).

129. Hooper, D. C. Mechanisms of fluoroquinolone resistance. *Drug Resist. Updat.* **2,** 38–55 (1999).

130. Kivisaar, M. Stationary phase mutagenesis : mechanisms that accelerate adaptation of microbial populations under. *Environmetal Microbiol.* **5,** 814–827 (2003).

131. Gill, W. P. *et al.* A replication clock for Mycobacterium tuberculosis. *Nat. Med.* **15,** 211–214 (2009).

132. Helaine, S. *et al.* Dynamics of Intracellular Bacterial Replication at the Single Cell Level. *Proc. Natl. Acad. Sci.* **107,** 3746–3751 (2010).

133. Ford, C. B. *et al.* Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. *Nat. Genet.* **43,** 482–6 (2011).

134. Obmolova, G., Ban, C., Hsieh, P. & Yang, W. Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* **407,** 703–710 (2000).

135. Fowler, R. G. *et al.* Interactions among the Escherichia coli mutT, mutM, and mutY damage prevention pathways. *DNA Repair (Amst).* **2,** 159–173 (2003).

136. Donnianni, R. A. & Symington, L. S. Break-induced replication occurs by conservative DNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 13475 (2013).

137. Witkin, E. M. Ultraviolet-Induced Mutation and DNA Repair. *Annu. Rev. Microbiol.* **23,** 487–514 (1969).

138. Witkin, E. M. . Radiation-Induced Mutations and Their Repair. *Science (80-.).* **152,** 1345–1353 (1966).

139. Anne Martin, H., Pedraza-Reyes, M., Yasbin, R. E. & Robelto, E. A. Transcriptional De-Repression and Mfd Are Mutagenic in Stressed Bacillus subtilis Cells. *J. Mol. Microbiol. Biotechnol.* 45–58 (2012). doi:10.1159/000332751

140. Pani, B. & Nudler, E. Mechanistic insights into transcription coupled DNA repair. *DNA*

Repair (Amst). **56,** 42–50 (2017).

141. Hayden, H. S. *et al.* Genomic Analysis of Salmonella enterica Serovar Typhimurium Characterizes Strain Diversity for Recent U.S. Salmonellosis Cases and Identifies Mutations Linked to Loss of Fitness under Nitrosative and Oxidative Stress. *MBio* **7,** 1–11 (2016).

142. Wolfgang, M. C. *et al.* Conservation of genome content and virulence determinants among clinical and environmental isolates of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci.* **100,** 8484–8489 (2003).

143. Witkin, E. M. Mutation and the Repair of Radiation Damage in Bacteria. *Radiat. Res. Suppl.* **6,** 30–53 (1966).

144. Richardson, A. R. *et al.* The Base Excision Repair System of Salmonella enterica serovar Typhimurium Counteracts DNA Damage by Host Nitric Oxide. *PloS Pathog.* **5,** (2009).

145. Schon, T. *et al.* Evaluation of wild-type MIC distributions as a tool for determination of clinical breakpoints for Mycobacterium tuberculosis. *J. Antimcrobial Chemother.* 786–793 (2009). doi:10.1093/jac/dkp262

146. Brandis, G., Pietsch, F., Alemayehu, R. & Hughes, D. Comprehensive phenotypic characterization of rifampicin resistance mutations in Salmonella provides insight into the evolution of resistance in Mycobacterium tuberculosis. *J. Antimicrob. Chemother.* **70,** 680–685 (2015).

147. Watson, M., Liu, J.-W. & Ollis, D. Directed evolution of trimethoprim resistance in Escherichia coli. *FEBS J.* **274,** 2661–2671 (2007).

148. Echols, H., Lu, C. & Burgers, P. M. Mutator strains of Escherichia coli, mutD and dnaQ, with defective exonucleolytic editing by DNA polymerase III holoenzyme. *Proc. Natl.*

Acad. Sci. U. S. A. **80,** 2189–2192 (1983).

149. Selby, C. P. & Sancar, A. Structure and function of transcription-repair coupling factor: I. Structural domains and binding properties. *Journal of Biological Chemistry* **270,** 4882–4889 (1995).

150. Branum, M. E., Reardon, J. T. & Sancar, A. DNA repair excision nuclease attacks undamaged DNA: A potential source of spontaneous mutations. *J. Biol. Chem.* **276,** 25421–25426 (2001).

151. Gao, Z., Wyman, M. J., Sella, G. & Przeworski, M. Interpreting the Dependence of Mutation Rates on Age and Time. *PLoS Biol.* 1–16 (2016). doi:10.1371/journal.pbio.1002355

152. Wistrand-Yuen, E. *et al.* Evolution of high-level resistance during low-level antibiotic exposure. *Nat. Commun.* **9,** (2018).

153. Blázquez, J. Hypermutation as a factor contributing to the acquisition of antimicrobial resistance. *Clin. Infect. Dis.* **37,** 1201–1209 (2003).

154. Cirz, R. T. *et al.* Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biol.* **3,** 1024–1033 (2005).

155. Cirz, R. T., Gingles, N. & Romesberg, F. E. Side effects may include evolution. *Nat. Med.* **12,** 890–891 (2006).

156. Mo, C. Y. *et al.* Systematically Altering Bacterial SOS Activity under Stress Reveals Therapeutic Strategies for Potentiating Antibiotics. *mSphere* **1,** 1–15 (2016).

157. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci.* **97,** 7090–7095 (2000).

158. Mayer, A., Landry, H. M. & Churchman, L. S. Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Curr. Opin. Cell Biol.* **46,** 72–80 (2017).

159. Chan, C. L., Wang, D. & Landick, R. Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3' end spacing distinguishes pause and termination pathways. *J. Mol. Biol.* **268,** 54–68 (1997).

160. Erie, D. A. The many conformational states of RNA polymerase elongation complexes and their roles in the regulation of transcription. *Biochim. Biophys. Acta* **1577,** 224–239 (2002).

161. Cheung, A. C. M. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471,** 249–253 (2011).

162. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. The RNA–DNA Hybrid Maintains the Register of Transcription by Preventing Backtracking of RNA Polymerase. *Cell* **89,** 33–41 (1997).

163. Nudler, E. RNA Polymerase Backtracking in Gene Regulation and Genome Instability. *Cell* **149,** 1438–1445 (2012).

164. Kang, J. Y. *et al.* RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol. Cell* **69,** 802–815.e1 (2018).

165. Kireeva, M. L. & Kashlev, M. Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc. Natl. Acad. Sci.* **106,** 8900–8905 (2009).

166. Tadigotla, V. R. *et al.* Thermodynamic and kinetic modeling of transcriptional pausing. *Proc. Natl. Acad. Sci.* **103,** 4439–4444 (2006).

167. Saba, J. *et al.* The elemental mechanism of transcriptional pausing. *Elife* **8,** (2019).

168. Winkler, M. E. & Yanofsky, C. Pausing of RNA polymerase during in vitro transcription of the tryptophan operon leader region. *Biochemistry* **20,** 3738–44 (1981).

169. Yakhnin, A. V., Yakhnin, H. & Babitzke, P. RNA Polymerase Pausing Regulates Translation Initiation by Providing Additional Time for TRAP-RNA Interaction. *Mol. Cell* **24,** 547–557 (2006).

170. Ray-Soni, A., Mooney, R. A. & Landick, R. Trigger loop dynamics can explain stimulation of intrinsic termination by bacterial RNA polymerase without terminator hairpin contact. *Proc. Natl. Acad. Sci.* **114,** E9233–E9242 (2017).

171. Turtola, M. & Belogurov, G. A. NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *Elife* **5,** 1–27 (2016).

172. Strauß, M. *et al.* Transcription is regulated by NusA:NusG interaction. *Nucleic Acids Res.* **44,** 5971–5982 (2016).

173. Selby, C. & Sancar, A. Molecular mechanism of transcription-repair coupling. *Science (80-.).* **260,** 53 (1993).

174. Christopher P. Selby. Mfd Protein and Transcription-Repair Coupling in E. coli. *Photochem Photobiol.* **93,** 280–295 (2017).

175. Zalieckas, J. M., Wray, L. V & Fisher, S. H. Expression of the Bacillus subtilis acsA gene: position and sequence context affect cre-mediated carbon catabolite repression. *J. Bacteriol.* **180,** 6649–54 (1998).

176. Westblade, L. F. *et al.* Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. *Nucleic Acids Res.* **38,** 8357–8369 (2010).

177. Prabha, S., Rao, D. N. & Nagaraja, V. Distinct Properties of Hexameric but Functionally Conserved Mycobacterium tuberculosis Transcription- Repair Coupling Factor. *PLoS One*

6, (2011).

178. Menoni, H., Hoeijmakers, J. H. J. & Vermeulen, W. Nucleotide excision repair-initiating proteins bind to oxidative DNA lesions in vivo. *J. Cell Biol.* **199,** 1037–1046 (2012).

179. McClure, W. R. & Cech, C. L. On the mechanism of rifampicin inhibition of RNA synthesis. *J. Biol. Chem.* **253,** 8949–8956 (1978).

180. Campbell, E. A. *et al.* Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* **104,** 901–912 (2001).

181. Nicolas, P. *et al.* Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science* **335,** 1103–6 (2012).

182. Mars, R. A. T., Nicolas, P., Denham, E. L. & van Dijl, J. M. Regulatory RNAs in Bacillus subtilis: a Gram-Positive Perspective on Bacterial RNA-Mediated Regulation of Gene Expression. *Microbiol. Mol. Biol. Rev.* **80,** 1029–1057 (2016).

183. Le, S. Y. & Maizel, J. V. A method for assessing the statistical significance of RNA folding. *J. Theor. Biol.* **138,** 495–510 (1989).

184. Freyhult, E., Gardner, P. P. & Moulton, V. A comparison of RNA folding measures. *BMC Bioinformatics* **6,** 241 (2005).

185. Borukhov, S., Polyakov, A., Nikiforov, V. & Goldfarb, A. GreA protein: A transcription elongation factor from Escherichia coli. *Proc. Natl. Acad. Sci.* **89,** 8899–8902 (1992).

186. Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12,** 87–98 (2011).

187. Brantl, S. & Jahn, N. SRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS Microbiol. Rev.* **39,** 413–427 (2015).

188. Dörr, T., Vulić, M. & Lewis, K. Ciprofloxacin Causes Persister Formation by Inducing the TisB toxin in Escherichia coli. *PLoS Biol.* **8,** e1000317 (2010).

189. Domka, J., Lee, J., Bansal, T. & Wood, T. K. Temporal gene-expression in Escherichia coli K-12 biofilms. *Environ. Microbiol.* **9,** 332–346 (2007).

190. Durand, S., Jahn, N., Condon, C. & Brantl, S. Type i toxin-antitoxin systems in Bacillus subtilis. *RNA Biol.* **9,** 1491–1497 (2012).

191. Silvaggi, J. M., Perkins, J. B. & Losick, R. Small Untranslated RNA Antitoxin in Bacillus subtilis. *J. Bacteriol.* **187,** 6641–6650 (2005).

192. Krüger, S., Gertz, S. & Hecker, M. Transcriptional analysis of bglPH expression in Bacillus subtilis: Evidence for two distinct pathways mediating carbon catabolite repression. *J. Bacteriol.* **178,** 2637–2644 (1996).

193. Irnov, I., Sharma, C. M., Vogel, J. & Winkler, W. C. Identification of regulatory RNAs in Bacillus subtilis. *Nucleic Acids Res.* **38,** 6637–6651 (2010).

194. Nilsson, R.-P., Beijer, L. & Rutberg, B. The glpT and glpQ genes of the glycerol regulon in Bacillus subtilis. *Microbiology* **140,** 723–730 (1994).

195. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches Control Fundamental Biochemical Pathways in Bacillus subtilis and Other Bacteria. *Cell* **113,** 577–586 (2003).

196. Ebbole, D. J. & Zalkin, H. Cloning and characterization of a 12-gene cluster from Bacillus subtilis encoding nine enzymes for de novo purine nucleotide synthesis. *J. Biol. Chem.* **262,** 8274–8287 (1987).

197. Eiamphungporn, W. & Helmann, J. D. Extracytoplasmic Function σ Factors Regulate Expression of the Bacillus subtilis yabE Gene via a cis-Acting Antisense RNA. *J.*

Bacteriol. **191,** 1101–1105 (2009).

198. Hung, S. C. & Gottesman, M. E. Phage HK022 Nun Protein Arrests Transcription on Phage lambda DNA in vitro and Competes with the Phage lambda N Antitermination Protein. *J. Mol. Biol.* 428–442 (1995).

199. Pathakt, V. K. & Temin, H. M. 5-Azacytidine and RNA Secondary Structure Increase the Retrovirus Mutation Rate. *J. Virol.* **66,** 3093–3100 (1992).

200. Lowe, T. M. *et al.* Transfer RNA genes experience exceptionally elevated mutation rates. *Proc. Natl. Acad. Sci.* **115,** 8996–9001 (2018).

201. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17,** 13 (2016).

202. Nicolas, P. *et al.* Condition-Dependent Transcriptome Architecture in Bacillus subtilis. *Science (80-.).* **1103,** 1103–1106 (2012).

203. Ragheb, M. N. *et al.* Inhibiting the Evolution of Antibiotic Resistance. *Mol. Cell* **73,** (2019).

204. Heewook, L., Ellen, P., Haixu, T. & Patricia, F. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci.* **109,** 1–6 (2012).

205. Kennedy, S. R. *et al.* Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat. Protoc.* **9,** 2586–2606 (2014).

206. Brehm, S. P., Staal, S. P. & Hoch, J. A. Phenotypes of pleiotropic-negative sporulation mutants of Bacillus subtilis. *J. Bacteriol.* **115,** 1063–70 (1973).

207. Harwood, C. ., Cutting, S. M. & Wiley, J. *Molecular Biological Methods for Bacillus*. (1990).

208. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **2000,** (2000).

209. Hoang, T. T., Karkho, R. R., Kutchma, A. J. & Schweizer, H. P. A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences : application for isolation of unmarked Pseudomonas aeruginosa mutants. *Gene* **212,** 77–86 (1998).

210. Kessel, J. C. Van & Hatfull, G. F. Recombineering in Mycobacterium tuberculosis. *Nat. Methods* **4,** 147–152 (2007).

211. Merrikh, H., Machón, C., Grainger, W. H., Grossman, A. D. & Soultanas, P. Co-directional replication-transcription conflicts lead to replication restart. *Nature* **470,** 554–7 (2011).

212. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

213. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

214. Homann, O. R. & Johnson, A. D. MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol.* **8,** 49 (2010).

215. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).

216. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

217. Hall, B. M., Ma, C., Liang, P. & Singh, K. K. Fluctuation AnaLysis CalculatOR : a web tool for the determination of mutation rate using Luria – Delbrück fluctuation analysis. *Bioinformatics* **25,** 1564–1565 (2009).

218. Deatherage, D. E. & Barrick, J. E. *Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq*. *Methods in Molecular Biology* **1151,** (2014).

219. Dove, S. L., Joung, J. K. & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein- protein contacts. *Nature* **386,** 627–630 (1997).

220. Weischenfeldt, J. & Porse, B. Bone marrow-derived macrophages (BMM): Isolation and applications. *Cold Spring Harb. Protoc.* **3,** 1–7 (2008).